

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)

Abstract

The fabled politics-administration dichotomy has largely been discredited as a guiding force in the field of public administration. The dichotomy lives on in the minds of many because nothing has emerged to take its place. Recent scholarship, however, has advanced the notion that the philosophy of American pragmatism could become such a force; some public administration scholars point to a so-called “implied pragmatism” as evidence. Implied pragmatism is the presence of pragmatic principles within the practice of public administration, even if not acknowledged as such. The present research explores this phenomenon within the state of Texas performance measurement system. This study reports the results of a survey that was distributed to all Texas state agencies and a subsequent focus group. Scant evidence of pragmatic principles was found within the dynamics of performance measure development. However, considerable evidence of pragmatism was found regarding the use and usefulness of performance measurement data. It was also found that small state agencies showed less evidence of pragmatic principles than did larger agencies.

Table of Contents

Chapter One: Introduction and Research Purpose.....	3 - 7
Chapter Two: Literature Review	
American Pragmatism.....	8 - 10
The Progressive Movement: Pragmatism Applied.....	11 - 12
Public Administration: No Pragmatists Need Apply.....	12 - 14
Performance Measurement: Pragmatism Implied?.....	14 - 16
Best Practices in Performance Measurement	16 - 17
Performance Measurement and Pragmatism.....	17 - 20
Constructing Performance Measures: The State of the Art and Science.....	20 - 24
Chapter Three: Setting	
Performance Measurement in Texas.....	25 - 29
State Auditor’s Office Survey	29 - 31
Discussion.....	31 - 32
Chapter Four: Conceptual Framework and Methodology.....	33 - 42
Chapter Five: Results	43 - 58
Chapter Six: Conclusions and Recommendations.....	59 - 66
Bibliography.....	67 - 71
Appendix A – Survey	72 - 74
Appendix B – Survey Data Frequencies Survey Data (Original).....	75 - 82
Appendix C – Agency Type Frequencies.....	83 - 84
Appendix D – Survey Data Frequencies (Recoded).....	85 - 92
Appendix E – Ratings and Values Legend.....	93 - 96

Chapter One: Introduction and Research Purpose

“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, 2000). 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.

What exactly is performance measurement? This applied research project will address the question in some detail. It would, however, be instructive at this point to offer the uninitiated reader some real world examples of performance measurement, and some of the underlying processes.

As a former Evaluation Analyst at the Texas Commission on Alcohol and Drug Abuse (TCADA), I was exposed to the promise and disappointment of performance measurement. TCADA contracts for two basic types of services: treatment and prevention of alcohol and drug abuse. Performance measurement of the former is quite successful; the latter, less so.

The Example of Alcohol and Drug Abuse Treatment

Treatment comes in three basic forms. Detoxification programs are designed to allow patients to rid themselves of toxic levels of alcohol and/or drugs in a quasi-medical

setting. Detoxification is not an end in itself, but a precursor to other forms of treatment. Accordingly, the measures for detoxification are the percentage of patients documented as ready for continued treatment, and the percentage of successful transfers to programs for additional treatment.

Drug-free treatment, either inpatient or outpatient, offers group and individual therapy, job counseling, relationship counseling, and the like. The goal is to provide patients with the requisite skills to function without the use of alcohol and drugs. Performance measures include the percentage of patients who completed treatment and were abstinent for 60 days after discharge.

Methadone programs provide heroin addicts with a legal, safe substitute drug in an outpatient setting. Patients can live a relatively normal life while treated with methadone. In this type of programming, the best results are obtained if the patient stays in the program for an extended period of time. Hence, performance is measured during treatment, rather than after discharge as above. Measures are the percentage of patients employed and abstinent during treatment, and the percentage of clients who are retained in treatment at least one year.

For each of the three treatment modalities, evaluation department staff developed measures after a careful review of the research literature. Draft measures were then circulated among other TCADA staff for review and comment. The measurement system was piloted with contracting treatment providers, allowing opportunities for comments and suggestions.

The Example of Alcohol and Drug Abuse Prevention Programs

Performance measurement of drug prevention programs is quite different from treatment programs, in part because prevention programming is much more diverse. Prevention programs tend to respond to the perceived specific needs of each community and can take a multitude of forms. Programming runs the gamut from fostering socialization of children just beginning school, to intervening when adolescents begin to display problem behaviors, to addressing college binge-drinking, to skill-building for parents of at-risk children, all the way to senior citizen prescription drug management, each with unique indicators of success. Treatment programs target alcohol and drug abusers; prevention programming has a much more heterogeneous target population. With this bewildering array of prevention approaches and target populations, it becomes extremely difficult to develop a standardized system to measure results as was done with treatment programs. What can be readily measured are the activities and services that prevention programs provide. That is, such things as attendance and enrollment can be readily counted; the actual results of the various programs cannot.

In the chapters that follow, 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 and extracted. Then, a method to test these hypotheses is developed and executed. Finally, results, conclusions, and recommendations are reported.

Chapter Two: Literature Review

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, 2000). 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, then, is based on the notions of practical experience and

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

workability as perceived at a given point in time. 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 (2000) 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. It is important to note that the community is involved throughout this experimental process, from developing working hypotheses through collecting, interpreting, and applying the resulting data. 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, 2000). So-called "experts" can also muffle other voices, inhibiting the free flow of ideas within a community of inquiry (Evans, 2000, p.319).

The Progressive Movement: Pragmatism Applied

Pragmatism provided the impetus for the Progressive Movement (Zanetti and Carr, 2000, p.433). Snider (2000a, p.333) credits pragmatism with providing the progressives with justification to challenge the existing social and political order. No longer were the appalling conditions and corrupt practices of the day accepted as part of the natural order²; pragmatism suffered no absolute truths.

As the Progressive Movement evolved from the agrarian populist campaigns of the late 1800s, its focus shifted to large urban areas (Caro, 1994 p.59). Many progressive reformers believed that corrupt politicians, in league with big business, were responsible for the squalid conditions in the big cities, and that government had the wherewithal to wrest control from these sinister forces. These reformers felt that government should enter heretofore unregulated areas of society, a rejection of the *laissez-faire* (Current et al., 1983, p.597 - 600).

Governments before the Progressive Movement were, as it seems now, startlingly uninformed. For instance, large cities did not have a budget as we know it. The total funding available at a given time was not known, nor was there a mechanism to determine the greatest civic needs so that funds could be allocated accordingly. The progressives felt that their first task was to scientifically analyze the existing order (Caro, 1994, p.60 - 61). With this approach, knowledge and expertise could be developed to intelligently and effectively lead (Current et al., 1984, p.600 – 603).

The progressives as a political force enjoyed a high degree of success. They spearheaded the passage of the 16th and 17th Amendments to the Constitution³ and numerous reform laws. Presidents Woodrow Wilson and Theodore Roosevelt took Progressivism to the highest office (M. Martin, 1978, p.512). The “ambitious social optimism” of the progressives eventually gave way to “pessimistic realism” in the post World War I period (Snider 2000a, p.337). Progressivism was no longer a force in American politics.

Public Administration: No Pragmatists Need Apply

The Progressive Movement came together around the time that the first writings regarding the recognizable field of public administration began to appear (Snider, 2000a, p.335). This self-awareness must have caused these writers to think about their budding field in the larger context of government and politics. What should be their place within this context? 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). Considering the prevailing notions of the time regarding politics and government, the founders of the field may have

² It is an irony of history that the Progressives sought to supplant the prevailing orthodoxy of Social Darwinism. Like pragmatism, Social Darwinism is a product of the revolution in thinking brought on by Darwin’s theory of evolution (Current et al., 1983, p.538). They are distant cousins.

chosen to disassociate themselves from the perceived corruption.

The origin of this separation, or dichotomy, as it came to be known, extends back to the writing of the Constitution. The traditional American skepticism about government (Radin, 1998, p.308) is institutionalized in the American governmental system. The fabled checks and balances among the executive, legislative, and judicial branches allow no single branch to hold sway over the others; this ensures freedom from tyranny. But the freedom gained through this separation of power comes at a price: inefficiency. The messy arena of governing pits branch against branch. The separation of administrators from politicians reflects the separation between the executive and legislative branches of government (Ellwood, 2000, p.324).

It would seem, at first glance, that public administration and pragmatism would make a companionable pairing. Indeed, many of the early writers saw themselves as pragmatists (Snider, 2000a, p.336). 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 writers⁴ 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).

³The 16th Amendment established a federal income tax. The 17th Amendment allowed for the direct popular

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). In fact, Svava (2001, p.176-177) maintains that the early public administration writers indeed saw a policy role for administrators, and further asserts that there has been and is now such a role. In spite of this, the notion of the dichotomy endures (Ingraham and Donahue, 2000, p.298). It is often still assumed, as a legacy of the dichotomy, that public administrators should have no voice in policy (Shields, 1996, p.394). Ellwood (2000, p.322) cites the absence of a guiding philosophy to replace the dichotomy as the primary reason for its persistence. Several contemporary scholars have suggested that pragmatism could assume a guiding role in the practice of public administration (Shields, 2000; Evans, 2000). The notion that pragmatic principles have crept into the practice of public administration has also been advanced.⁵ Performance measurement is one area where there seems to be significant implied pragmatism (Shields, 2000); this idea will be addressed in detail below.

Performance Measurement: Pragmatism Implied?

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

election of Senators; they were elected by state legislatures prior to this amendment.

⁴ Notably, Frank Goodnow and Charles A. Beard.

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).

The recent stimulus for the rise of performance measurement comes from taxpayer unrest and from Osborne and Gaebler's *Reinventing Government* (1992). This influential book in turn spawned a spate of legislation aimed at integrating performance measurement into the every-day business of government (Joyce, 1993, p.1-2). The impetus for performance measurement is not new, however. It dates back at least 50 years, to a time when the first efforts at rationalizing the budget process were made. Since then, there have been several, largely unsuccessful attempts at reforming the process.⁶ Management by Objective (Osborne and Gaebler, 1992, p.156-159), Total Quality Management, and the advent of managed care added to this impetus along the way (Martin and Kettner, 1997, p.18-19).

Another precursor to performance measurement is program evaluation. Program evaluation is designed to gather the same sort of information as performance measurement but typically is more rigorous and formal (Allen, 1996, p.13). Ideally,

⁵ Shields, 1996; Stever, 2000; and Svava, 2001.

program evaluation is experimentation carefully designed to determine the relationship between interventions and the predicated outcomes (Mullen and Magnabosco, 1997, p.4). Program evaluation thrived in the 1960s as skepticism regarding social programs began to rise and costs to collect and process data began to decrease (Rossi, 2000, p.20). Performance measurement is less expensive than program evaluation, but is not typically conclusive about cause and effect (Mullen and Magnabosco, 1997, p.4).

Best Practices in Performance Measurement

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 to determine allocations for the upcoming fiscal year, known as performance funding (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 allows managers to monitor results and progress on a continuing basis, and offers staff clear direction regarding the most important objectives. Forrester and Adams (1997, p.475-476) urge that performance measurement serve both budgetary and management functions.

⁶ For a complete discussion of historical federal budget reform efforts, see *Using Performance Measures in*

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, and more importantly, outcomes, the measurable results of these 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.⁷

It is also important to construct a measurement system that accurately represents 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.

the Federal Budget Process (GAO, 1993).

⁷ Note that in the examples of performance measurement in the introduction, alcohol and drug abuse treatment programs measured outcomes while prevention programs measured outputs.

Performance Measurement and Pragmatism

Shields (2000, p.97) has suggested that pragmatic principles could be readily applied to the performance measure process. 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.⁸

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 best practices is the notion of putting theory to the practical test. Performance measurement can be viewed as an inquiry into a particular program or intervention. Government programs have an overall purpose. 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 about outcome can then

⁸ 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).

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. As noted above, performance measurement seldom includes the controls that typically attend program evaluation; nonetheless, care should be exercised when measures are defined.

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.

The commonalities noted above suggest that pragmatism could be readily applied

to the performance measurement process. Stever (2000, p.455) proposes that pragmatism is implied in some public administration activities. Others seem to support this thesis as well. Shields (1996) posits that pragmatism explains public administration's effect on policy. Svara (2001, p.180), in debunking the politics-administration dichotomy, refers to the concept of "complementarity", which "entails ongoing interaction, reciprocal influence, and mutual deference between elected officials and administrators.

Administrators help to shape policy, and they give it specific content and meaning in the process of implementation. Elected officials oversee implementation, provide specific complaints about poor performance, and attempt to correct problems with performance through fine-tuning." This notion sounds very much akin to the community of inquiry concept.

Constructing Performance Measures: The State of the Art and Science

Performance measurement is here to stay. Over the last 10 years, it has become an essential management tool (Ingraham, 2000, p.303). However, 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 affect

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, either outputs, outcomes, or neither can be measured owing to the types of activities and services provided. 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.⁹

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.¹⁰ 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, agencies that do not provide direct services, but instead contract for same, will be one step removed from these services and activities; they are often held accountable for outcomes not entirely under their control (Joyce, 1993, p.10). 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. Yet this failure is often cited as a reason for the ineffectiveness of previous budgeting reform efforts (GAO, 1993, p.26).

⁹ This is another reason why the outcomes were not measured in the prevention programming example offered in the introduction.

¹⁰ Note that these factors are the same as threats to internal validity in experimental design. That is, an alternate explanation cannot be ruled out.

Kravchuk and Shack (1999, p.353-354) maintain that an organization's proximity to direct service provision affects the type of outcome measures it will favor. Agencies that provide direct services are likely to construct measures in such a way as to defend or increase their funding. Organizations that contract out for direct services are likely to promote effectiveness measures as a way to monitor and manage their contractors. The parties involved with a particular endeavor may have very different views of what should be measured, depending on their role and function in the system; this has important implications for a community of inquiry.

Working With Data

In addition to the difficulty in constructing appropriate performance measures, the mechanics of measure development and performance data use 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. A 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 on achieving these simple objectives at the expense of other important functions that are not represented in the array of measures.

The link between performance measurement and budgeting for the purposes of agency accountability can engender a certain fear in program managers (Broom, 1995, p.14). In response to this fear, managers may construct measures that are easily accomplished, falsify data, or lowball targeted levels of performance (GAO, 1993, p.26; Heinrich, 1999, p.371).

Finally, agencies may be hampered by a limited or antiquated management information system (MIS). Radin (1999 p.312) found a resistance to upgrading MIS at the federal level. The state of Maine delayed their ambitious performance-based budgeting plans, at least in part, because of an antiquated MIS (Harris, 1999).

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 chapter addresses the setting of this study.

Chapter Three: Setting

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. This chapter describes 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).

Strategic plans in Texas cover a five-year period and identify each agency's mission and goals, performance measures and projected performance targets, target population, and an analysis of the funding needed to accomplish the stated mission and goals. Agencies are required to develop and publish their strategic plan each odd numbered year (LBB, 1999a). Strategic planning confers many benefits to agencies (LBB, 2000):

- clarifies purpose and direction
- helps develop a strong identity
- enhances decision-making
- strengthens internal and external communication
- articulates the primary factors affecting the agency
- guides budget preparation

As noted above, a major component of the strategic plan is the development of agency performance measures. Measures are established for a two-year period, with annual performance targets over that biennium. Performance measures come in four varieties in Texas. Output and outcome measures have been defined previously. Efficiency measures are designed to determine unit cost of services or items. Explanatory measures are designed to provide additional information believed to affect other measures, and usually take the form of outputs. Table 3.1 below displays examples of

each type of measure surrounding adult alcohol and drug abuse treatment.¹¹ Note that a comprehensive strategic plan, with associated measures in all their forms, should provide an overall description of agency functioning.

Table 3.1 State of Texas Measure Definitions, Examples, and Targets

Measure Type	Definition	Example	FY02 Target	FY03 Target
Outcome	indicates the actual effect upon a stated condition	Percent of adults entering treatment programs who report that they are abstinent when contacted following discharge	81%	81%
Output	counts the services produced by an agency	Number of adults served in treatment programs	33,699	33,699
Efficiency	gauges resource cost per unit of product	Average cost per adult in treatment programs	\$1919	\$1919
Explanatory	provides information to help assess reported performance	Percent of adults completing treatment programs	71%	71%

As part of the strategic plan, agencies must also develop measure definitions. Definitions include a narrative explanation of the measure, the method of calculation, and a description of outside influences that may affect the measure’s validity. The State Auditor’s Office (SAO) conducts on-site reviews of performance data systems to verify accuracy and determine how agency management uses this data. Performance measure achievement is reported quarterly and annually. A variation of 5% between projected and actual targeted performance must be explained (LBB, 1999a). Performance can be rewarded and punished in Texas. Table 3.2 below displays the reward and penalty options available to the legislature (LBB, 1999b, p.17).

¹¹ These measures are taken verbatim from the General Appropriations Act 2002 – 2003. Available at: http://www.lbb.state.tx.us/The_LBB/Access/AppBills_LBEs.htm

Table 3.2 State of Texas Performance Rewards and Penalties

<p>Rewards</p>	<ul style="list-style-type: none"> • Increased funding • Exemption from reporting requirements • Increased funding transferability • Formalized recognition or accolade • Awards or bonuses • Expanded responsibility • Expanded contracting authority
<p>Penalties</p>	<ul style="list-style-type: none"> • Reduction of funding • Elimination of funding • Restriction of funding • Withholding of funding • Reduction of transferability • Transfer of functional responsibility to another entity • Recommendation for placement in conservatorship • Direction that a management audit be conducted • Direction that other remedial or corrective action plans be implemented.

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. That is to say, performance measure data must be used in the day-to-day decision making process. The LBB must approve measures and revisions, but agency management is responsible for their development (LBB, 1999b. p.7-8). It is expected that a significant and frequent dialogue be established between the LBB and agency management staff.

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.¹² 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. Forty-four (44%) percent of the respondents believed that performance measures should be eliminated from the General Appropriations Act. Only 43% felt that the link between performance and budget "always" or "almost always" increases accountability.
3. Less than half (47%) of the respondents believed that performance measure data was "always" or "almost always" used for agency management.
4. When asked if the key measures represent the most important indicators of agency performance, less than half of the respondents (46%) agreed.
5. Only one-third (34%) of the respondents believed that performance

measures “always” or “almost always” provided an early warning system for problems in their agency.

6. Sixty-nine percent (69%) of the respondents did not think that rewards and penalties should be tied to agency performance.
7. Only 21% of the respondents said they “always” or “almost always” include performance measures in their subcontracts.
8. When asked to list three 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 suggested improvement 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.¹³ 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. The differential ranged from 2% to 28%.¹⁴

It has been argued earlier in this paper that many tenets of pragmatism are present

¹² Overall, 176 state agencies responded to the survey. About one-third were education agencies; another fifth were regulatory.

¹³ A large agency’s total budget is greater than or equal to \$40 million; a small agency is less than \$40 million.

in the best practice of performance measure development. It is interesting to note that this is also true of the Texas performance measure process. Commonalities include the development of a comprehensive array of measures in the strategic plan, inclusive participation in measure definition, and the day-to-day practical use of data.

Discussion

The philosophy of pragmatism and the field of public administration never mingled, at least openly, but it seems that they should have (Evans, 2000). Recent scholarship has effectively made the point that pragmatism has much to offer to public administration¹⁵, but the early establishment of the politics-administration dichotomy veered the field, at least in theory, away from pragmatism.

Recent scholarship has also suggested that some pragmatic principles have anonymously crept into the practice, if not the theory, of public administration.¹⁶ There does not, however, seem to be consensus regarding the extent of this “implied pragmatism.” (Stever, 2000, p.455-456).

Performance Measurement as Practiced in Texas

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

¹⁴ The nine performance measure related questions are numbers: 1, 4, 6, 7, 8, 9, 10, 12, and 17. The way the data were reported by SAO makes it difficult to determine percentages for some questions. See (SAO, 1998).

¹⁵ Evans (2000), Shields (1996), Snider (2000)

literature indicates that there remain significant problems in the practice of measure development. Practitioners feel that there are substantial problems with the performance measures in use.

An analogous situation seems to have occurred in Texas as well. 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. A significant portion of the respondents to the SAO survey felt the quality of measures, and the process to achieve measure definition, were problematic.

In the discussion so far, the principles of American pragmatism have been linked to the best practices of performance measurement. Some of the pitfalls of actual practices have been addressed as well. What remains is to construct a system of inquiry into performance measurement as practiced in Texas. That is the subject of the next chapter.

¹⁶ Shields (1996), Stever (2000); Svava (2001)

Chapter Four: 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 chapter, 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

As noted above, several scholars have recently advanced the idea that the philosophy of pragmatism may have much to offer the field of public administration. 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. 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. The hypotheses are based on the premise that in order for measures to be well constructed, in light of the many pitfalls associated with measure design, scientific principles must have been applied. The first working hypothesis centers around the use of an ongoing, iterative, and experimental process in measure development. A second 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.

Table 4.1 Working Hypotheses

The Performance Measure Development Process			
WH1	Literature Sources	The Texas state agency performance measure process employs “community of inquiry” principles.	Research Methods
WH1a	Shields (2000) Ellwood (2000)	The measure development process involves active input from a broad array of sources.	Survey Q1 & 2 Focus Group
WH1b	Shields (2000) Evans (2000)	Participants are mutually responsive.	Survey Q3-Q6 Focus Group
The Use of Scientific Principles			
WH2	Literature Sources	Sound scientific practices are employed in performance measure use and development.	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
The Use of Performance Measurement Data.			
WH3	Literature Sources	Performance data are used to guide agency operations.	Research Methods
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
Large versus small agencies			
WH4	Literature Sources	Pragmatic principles are less evident in small state agencies compared with large agencies.	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

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 (Babbie, 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; 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 survey¹⁷ was developed by the author and was pretested by two former state agency employees. The unit of analysis for the survey data is the state agency; one or more staff responded to the survey on behalf of the 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
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.

Survey Recoding

In the analysis results reported in Chapter 5, 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.¹⁸

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. Only four of the invited respondents agreed to participate, one of whom did not attend because of illness.

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

¹⁷ The survey is located in Appendix A

¹⁸ Frequency tables representing the original coding are found in Appendix B.

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 webpage¹⁹ as of September 23, 2001. Quasi-state agencies²⁰ which appeared on the webpage were excluded, as was the Legislative Budget Board for obvious reasons. Surveys were sent to all other listees. One survey was returned incomplete noting that the agency had been eliminated at the beginning of the fiscal year.²¹ Another survey was returned incomplete because the agency was housed in the legislative branch and did not participate in the performance measurement process.²² A third was discarded when handwritten comments indicated that the respondent did not understand what was being asked. These three surveys were excluded from study calculations.

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 for a response rate of 40%.

¹⁹ <http://www.tsl.state.tx.us/trail/agencies.html>

²⁰ Lower Colorado River Authority, Texas Guaranteed Student Loan

²¹ Finance Commission

²² Sunset Commission

Analysis of Sample

A chi-square analysis 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. Agency type was self-reported on the survey whereas, for the overall sample, values for this variable were determined from the 2002-2003 Appropriations Act.²³ It should be noted that in several instances, agencies were not sure how to code this item on the survey and indicated two agency types. In these cases, the data were thrown out for this variable.

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, $\chi^2(6, n = 81) = 30.08, p. < .00005$. For this reason, any inferences regarding the state of Texas performance measurement system as a whole should be made only with extreme caution. Chi-square results are shown in Table 4.2 below.²⁴

Note that observed frequencies for Education and Health & Human Services agencies are substantially higher than expected frequencies, with residual values of 10.57 and 6.19, respectively. Conversely, the absolute residual value for Universities was nearly nine.

²³ Available at: http://www.lbb.state.tx.us/The_LBB/Access/AppBills_LBEs.htm

²⁴ Agency type frequencies, without the combining of categories used in the chi-square, are found in Appendix C.

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

Survey respondents were asked to categorize their agencies as large or small based on whether their total budget exceeded \$40 million or not. As shown in Table 4.3 below, just over half (52.6%) that answered this question were large agencies.

Table 4.3 Responding agencies by budget size

	Frequency	Percent
Budget greater than or equal to \$40 million	45	53.6
Budget less than \$40 million	39	46.4
Total	84	100.0

The study was executed as described above during September of 2001. Results are presented and discussed in Chapter 5 below.

Chapter Five: 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 is 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 is presented and discussed along with each of survey questions. That is to say, there is no separate section for reporting results of the fourth working hypothesis.

The data tables, as noted in the previous chapter, contain the percentage of respondents who combine to “Agree” or “Strongly agree” with each survey question. 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 will be 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.²⁵ 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.

Another focus group participant cited an incipient effort within the Health and Human Services Commission (HHSC) that may be a sign of things to come, and has significant community of inquiry implications. HHSC is an umbrella agency that has responsibility over the 12 health and human services agencies in Texas. In an attempt to coordinate similar programmatic efforts, HHSC has fostered the development of cross-agency, system-wide measures with multiple agency involvement. If this effort comes to fruition, it could become a major component in forming a health and human services community of inquiry.

²⁵ See Appendix D for a table showing “Combined Agree”, “Combined Disagree” and “Neutral” percentages for each question (Table Q1).

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

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%.

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 disagree²⁶, 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.

²⁶ See Appendix D for a table showing “Combined Agree”, “Combined Disagree” and “Neutral” percentages for each question (Table Q3).

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

As noted in Chapter 3, 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 indicates, 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.

The above analysis indicates the respondents' view that the power in the relationship between the LBB and state agencies is tilted toward the LBB. This analysis,

of course, includes those who agreed, disagreed, and were neutral regarding the premises in these questions. The data is even more telling when considering these same three survey items for a subset of respondents who disagreed or strongly disagreed with the notion that the relationship between the principal parties was equal (n = 46). Data from this subset of respondents shows even less agreement regarding measure initiation, modification, and elimination (see Table 5.3 below). This subset of respondents is a striking 18 percentage points less likely to agree with the notion that most performance measures are initiated by agency staff. Percentages are about nine points lower regarding LBB agreement to modify or eliminate measures.

Table 5.3 Comparison of All Respondents with Subset

	All % Agree	Subset % Agree	Difference
Measures are initiated by agency staff.	38.1	20.0	18.1
LBB agrees to modify measure definitions.	42.7	33.3	9.4
LBB agrees to eliminate measures.	32.5	23.9	8.6

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.

While the focus group's view was more charitable than the data seemed to indicate, the participants did cite instances where measures were forced upon state agencies, often at the behest of a legislator.

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

Discussion

There is only weak support among the respondents for the overall hypothesis that the state of Texas employs community of inquiry principles in the performance measurement development process. Agencies do not typically solicit input from outside stakeholders for program measure development. The power in the measure development process seems to reside with the LBB, rather than being shared in an equal, mutually responsive arrangement. Agencies do, however, seem to involve their own staff in the measure development process.

The findings for small agencies are, across the board, even less supportive of these sub-hypotheses than larger agencies; in every case, small agencies had a smaller combined agree score. In some cases, the difference was striking.

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 this 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. It is noted that small agency support drops below the 50% criterion (42.1%) regarding the proposition that performance measures change as agency functions change.

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 items²⁷ that follow. Just over 7% of the respondents felt there were too few measures. Nearly 39%

²⁷ Note that the 50% decision rule does not apply to explanatory items.

thought there were too many, but a substantial minority of 31.8% combined to disagree.²⁸

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

There is an interesting anomaly when comparing data associated with Working Hypothesis 1b with the present hypothesis. Table 5.2 above indicates that only 32.5% of the respondents agree with the assertion that the LBB allows agency requests to eliminate performance measures. This would lead, one might suppose, to a number of superfluous measures. However, the data in Table 5.6 indicates that the respondents, on the whole, don't feel there are too many measures.

The focus group and a survey respondent reported that agencies sometimes have an alternate performance measurement system in place. These agencies gather what they

²⁸ See Appendix D for a table showing “Combined Agree”, “Combined Disagree” and “Neutral” percentages for each question (Table Q12).

consider to be useful information in addition to LBB measures. 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 indicates 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

Discussion

The survey respondents provided considerable support for the presence of scientific principles in the Texas performance measurement process. At least half thought that measure development process responded to changing needs, and that data were used to modify measures and targets. Similarly, a majority of respondents reported that measures are accurate representations of agency functions. Agency staff was generally perceived to be well able to effectively use performance data. While these “micro” processes seem to be working quite well, the “macro” process would seem to be flawed. Only a third of the respondents supported the notion that the measures in combination accurately represented overall agency functioning. It should be noted, as discussed earlier in this presentation, that some functions and activities are difficult to measure by their very nature; this phenomenon might be responsible for the agencies’ lack of support for this item.

Small agencies are generally less supportive of the hypotheses regarding the use of scientific principles in the performance measurement process. This is true for each non-explanatory survey item with one exception: regarding agency staffs' ability to apply performance measure data, small and large agency percentages were virtually equal.

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.

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 did large agencies. 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.²⁹ 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, with by far the smallest number of responses (n = 60). Nearly equal agreement and disagreement percentages with a high number of neutral responses (value = 3) pull the mean (3.0) toward the benchmark. It seems, and the focus group concurred, that this question may not have been well understood by the respondents. Small agency agreement with this hypothesis was very low, almost 27 percentage points below the large agency rate. See Table 5.8, below.

Discussion

The hypothesis that data are put to optimal use by state agencies has at least mixed support. At least half the respondents combine to agree that performance data are used to guide allocation of funds and to improve agency operations. Small agencies agreed with these premises at a higher rate than larger agencies. The proposition that data are used to improve contractor performance does not seem to be supported, although the

²⁹See Appendix D for a table showing “Combined Agree”, “Combined Disagree” and “Neutral” percentages for each question (Table Q19).

veracity of this data may be called into question.

The focus group, while acknowledging that data are used to modify agency function, said that the most powerful and compelling force for change are audits by the Sunset Commission.

Table 5.8 Use of performance data to effect 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

The data above are organized and reported by sub-hypothesis. In the following chapter, a method of operationalizing the findings for each sub-hypothesis and hypothesis is employed to determine levels of support for each.

Chapter Six: 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). See Appendix E for details on coding, ratings and calculations.

Table 6.1 Sub-hypothesis Support Ratings

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 illustrated graphically 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 was not supported at all. 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.

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	

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 is 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 completely 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 6.3 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, and in some cases, vocal 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.

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 (see Chapter 2 above). An inquiry into this aspect of performance measurement might enrich the field.

4. Several participants in this study, both verbally and as comments on surveys, expressed some fear that the LBB would be able to attribute their responses back to them. Further, there is a substantial difference in the survey response rate between email (34%) and regular mail (45.3%) respondents. Mail respondents could be anonymous; email respondents could not. This differential may indicate a fear of being identified. Finally, there may have been some trepidation in attending a focus group where free expression of ideas would be expected. One attendee quizzed me very carefully on how focus group data would be used, and the confidentiality he could expect. It is quite possible, considering the above, that some parties were reluctant to respond out of fears regarding a possible breach of confidentiality. If this speculation is correct, it would have serious community of inquiry implications. Of course, further study into this aspect would likely encounter the same apprehension.

It might be advisable to conduct further studies into performance measurement with the approval and blessing of the LBB, with full confidentiality guaranteed.

5. The focus group, in their closing remarks, made the point that allocation of funds to Texas state agencies remains essentially a political process, despite one of the best and most recognized performance measurement systems in the country. The group went so far as to recommend that performance data be used for management purposes only. That is, the link between state agency funding and performance should be severed. An inquiry

into the separation of performance measure from the budget process might prove interesting.

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.

Harris, J. L. (1999). Performance budgeting is on the move, PA Times, 22(8), p.1-2.

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.

King, C. S. & Stivers, C. (1998). Government Is Us: Public Administration in an Anti-Government Era. Thousand Oaks, CA: Sage Publications.

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 (pp.199-230).

Shields, P. M. (2000). The community of inquiry: Classical Pragmatism and Public Administration. Manuscript submitted for publication.

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

Appendix B

Q1

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
Strongly Agree	5	6	7.0	7.1	7.1
Agree	4	22	25.6	26.2	33.3
Neutral	3	24	27.9	28.6	61.9
Disagree	2	21	24.4	25.0	86.9
Strongly Disagree	1	11	12.8	13.1	100.0
.		2	2.3	Missing	
		-----	-----	-----	
Total		86	100.0	100.0	

Valid cases 84 Missing cases 2

Q2

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
Strongly Agree	5	14	16.3	16.5	16.5
Agree	4	44	51.2	51.8	68.2
Neutral	3	6	7.0	7.1	75.3
Disagree	2	16	18.6	18.8	94.1
Strongly Disagree	1	5	5.8	5.9	100.0
.		1	1.2	Missing	
		-----	-----	-----	
Total		86	100.0	100.0	

Valid cases 85 Missing cases 1

Q3

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
Strongly Agree	5	2	2.3	2.4	2.4
Agree	4	14	16.3	17.1	19.5
Neutral	3	20	23.3	24.4	43.9
Disagree	2	31	36.0	37.8	81.7
Strongly Disagree	1	15	17.4	18.3	100.0
.		4	4.7	Missing	
		-----	-----	-----	
Total		86	100.0	100.0	

Valid cases 82 Missing cases 4

Appendix B

Q4

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
Strongly Agree	5	6	7.0	7.1	7.1
Agree	4	26	30.2	31.0	38.1
Neutral	3	14	16.3	16.7	54.8
Disagree	2	27	31.4	32.1	86.9
Strongly Disagree	1	11	12.8	13.1	100.0
.	.	2	2.3	Missing	
Total		86	100.0	100.0	

Valid cases 84 Missing cases 2

Q5

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
Strongly Agree	5	4	4.7	4.9	4.9
Agree	4	31	36.0	37.8	42.7
Neutral	3	22	25.6	26.8	69.5
Disagree	2	17	19.8	20.7	90.2
Strongly Disagree	1	8	9.3	9.8	100.0
.	.	4	4.7	Missing	
Total		86	100.0	100.0	

Valid cases 82 Missing cases 4

Q6

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
Strongly Agree	5	1	1.2	1.2	1.2
Agree	4	26	30.2	31.3	32.5
Neutral	3	26	30.2	31.3	63.9
Disagree	2	25	29.1	30.1	94.0
Strongly Disagree	1	5	5.8	6.0	100.0
.	.	3	3.5	Missing	
Total		86	100.0	100.0	

Valid cases 83 Missing cases 3

Appendix B

Q7

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
Strongly Agree	5	10	11.6	12.2	12.2
Agree	4	31	36.0	37.8	50.0
Neutral	3	24	27.9	29.3	79.3
Disagree	2	14	16.3	17.1	96.3
Strongly Disagree	1	3	3.5	3.7	100.0
.	.	4	4.7	Missing	
Total		86	100.0	100.0	

Valid cases 82 Missing cases 4

Q8

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
Strongly Agree	5	10	11.6	12.0	12.0
Agree	4	37	43.0	44.6	56.6
Neutral	3	14	16.3	16.9	73.5
Disagree	2	20	23.3	24.1	97.6
Strongly Disagree	1	2	2.3	2.4	100.0
.	.	3	3.5	Missing	
Total		86	100.0	100.0	

Valid cases 83 Missing cases 3

Q9

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
Strongly Agree	5	21	24.4	24.7	24.7
Agree	4	44	51.2	51.8	76.5
Neutral	3	11	12.8	12.9	89.4
Agree	2	8	9.3	9.4	98.8
Strongly Disagree	1	1	1.2	1.2	100.0
.	.	1	1.2	Missing	
Total		86	100.0	100.0	

Valid cases 85 Missing cases 1

Appendix B

Q10

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
Strongly Agree	5	6	7.0	7.1	7.1
Agree	4	37	43.0	43.5	50.6
Neutral	3	20	23.3	23.5	74.1
Disagree	2	18	20.9	21.2	95.3
Strongly Disagree	1	4	4.7	4.7	100.0
.	.	1	1.2	Missing	
Total		86	100.0	100.0	

Valid cases 85 Missing cases 1

Q11

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
Strongly Disagree	5	6	7.0	7.1	7.1
Agree	4	22	25.6	25.9	32.9
Neutral	3	21	24.4	24.7	57.6
Disagree	2	29	33.7	34.1	91.8
Strongly Disagree	1	7	8.1	8.2	100.0
.	.	1	1.2	Missing	
Total		86	100.0	100.0	

Valid cases 85 Missing cases 1

Q12

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
Strongly Agree	5	11	12.8	12.9	12.9
Agree	4	22	25.6	25.9	38.8
Neutral	3	25	29.1	29.4	68.2
Disagree	2	26	30.2	30.6	98.8
Strongly Disagree	1	1	1.2	1.2	100.0
.	.	1	1.2	Missing	
Total		86	100.0	100.0	

Valid cases 85 Missing cases 1

Appendix B

Q13

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
Agree	4	6	7.0	7.1	7.1
Neutral	3	25	29.1	29.4	36.5
Disagree	2	41	47.7	48.2	84.7
Strongly Disagree	1	13	15.1	15.3	100.0
.	.	1	1.2	Missing	
Total		86	100.0	100.0	

Valid cases 85 Missing cases 1

Q14

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
Strongly Agree	5	19	22.1	22.4	22.4
Agree	4	46	53.5	54.1	76.5
Neutral	3	10	11.6	11.8	88.2
Disagree	2	9	10.5	10.6	98.8
Strongly Disagree	1	1	1.2	1.2	100.0
.	.	1	1.2	Missing	
Total		86	100.0	100.0	

Valid cases 85 Missing cases 1

Q15

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
Strongly Agree	5	22	25.6	25.9	25.9
Agree	4	54	62.8	63.5	89.4
Neutral	3	6	7.0	7.1	96.5
Disagree	2	2	2.3	2.4	98.8
Strongly Disagree	1	1	1.2	1.2	100.0
.	.	1	1.2	Missing	
Total		86	100.0	100.0	

Valid cases 85 Missing cases 1

Appendix B

Q16

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
Strongly Agree	5	18	20.9	21.4	21.4
Agree	4	54	62.8	64.3	85.7
Neutral	3	7	8.1	8.3	94.0
Disagree	2	4	4.7	4.8	98.8
Strongly Disagree	1	1	1.2	1.2	100.0
.	.	2	2.3	Missing	
Total		86	100.0	100.0	

Valid cases 84 Missing cases 2

Q17

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
Strongly Agree	5	8	9.3	9.8	9.8
Agree	4	31	36.0	37.8	47.6
Neutral	3	19	22.1	23.2	70.7
Disagree	2	18	20.9	22.0	92.7
Strongly Disagree	1	2	2.3	2.4	95.1
Not Applicable	0	4	4.7	4.9	100.0
.	.	4	4.7	Missing	
Total		86	100.0	100.0	

Valid cases 82 Missing cases 4

Q18

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
Strongly Agree	5	9	10.5	10.6	10.6
Agree	4	41	47.7	48.2	58.8
Neutral	3	14	16.3	16.5	75.3
Disagree	2	16	18.6	18.8	94.1
Strongly Disagree	1	5	5.8	5.9	100.0
.	.	1	1.2	Missing	
Total		86	100.0	100.0	

Valid cases 85 Missing cases 1

Appendix B

Q19

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
Strongly Agree	5	2	2.3	2.5	2.5
Agree	4	14	16.3	17.7	20.3
Neutral	3	27	31.4	34.2	54.4
Disagree	2	16	18.6	20.3	74.7
Strongly Disagree	1	1	1.2	1.3	75.9
Not Applicable	0	19	22.1	24.1	100.0
	.	7	8.1	Missing	
	Total	86	100.0	100.0	
Valid cases	79	Missing cases	7		

Appendix C

Appendix C

Q21 Agency Type					
Value Label	Frequency	Percent	Valid Percent	Cum Percent	
General Government	6	7	7.4	7.4	
Education	17	19.8	21	28.4	
Medical Institution	3	3.5	3.7	32.1	
Public Safety	6	7	7.4	39.5	
Business & Econ. Dev	2	2.3	2.5	42	
Health and Human Svcs.	13	15.1	16	58	
University	10	11.6	12.3	70.4	
Judiciary	4	4.7	4.9	75.3	
Natural Resources	3	3.5	3.7	79	
Regulatory	17	19.8	21	100	
Missing	5	5.8	---	---	
	Total	86	100	100	
Valid cases	81	Missing cases	5		

Appendix D

Q1- My agency solicits input from interested outside parties in developing performance measures.(e.g. contractors, community)

	Frequency	Percent	Valid Percent	Greater than \$40M	Less than \$40M	Difference
Combined Agree	28	32.6	33.3	40.9	23.7	17.2
Neutral	24	27.9	28.6	25.0	34.2	-9.2
Combined Disagree	32	37.2	38.1	34.1	42.1	-8.0
Missing	2	2.3	---	---	---	---
Total	86	100.0	100.0	100.0	100.0	
Valid cases	84	Mean	2.89			

Q2 - Agency line staff are involved in performance measure development.

	Frequency	Percent	Valid Percent	Greater than \$40M	Less than \$40M	Difference
Combined Agree	58	67.4	68.2	72.7	61.5	11.2
Neutral	6	7.0	7.1	4.5	10.3	-5.8
Combined Disagree	21	24.4	24.7	22.7	28.2	-5.5
Missing	1	1.2	---	---	---	---
Total	86	100.0	100.0	99.9	100.0	
Valid cases	85	Mean	3.54			

Q3 - Agency staff and LBB staff are equals in performance measure development.

	Frequency	Percent	Valid Percent	Greater than \$40M	Less than \$40M	Difference
Combined Agree	16	18.6	19.5	20.9	13.5	7.4
Neutral	20	23.3	24.4	23.3	27.0	-3.7
Combined Disagree	46	53.5	56.1	55.8	59.5	-3.7
Missing	4	4.7	---	---	---	---
Total	86	100.0	100.0	100.0	100.0	
Valid cases	82	Mean	2.48			

Q4 - Most performance measures are initiated by agency staff rather than LBB staff.

	Frequency	Percent	Valid Percent	Greater than \$40M	Less than \$40M	Difference
Combined Agree	32	37.2	38.1	37.8	35.1	2.7
Neutral	14	16.3	16.7	15.6	18.9	-3.3
Combined Disagree	38	44.2	45.2	46.7	45.9	0.8
Missing	2	2.3	---	---	---	---
Total	86	100.0	100.0	100.1	99.9	
Valid cases	84	Mean	2.87			

Q5 - LBB staff agree to agency requests to modify performance measure definitions.

	Frequency	Percent	Valid Percent	Greater than \$40M	Less than \$40M	Difference
Combined Agree	35	40.7	42.7	48.8	35.1	13.7
Neutral	22	25.6	26.8	27.9	24.3	3.6
Combined Disagree	25	29.1	30.5	23.3	40.5	-17.2
Missing	4	4.7	---	---	---	---
Total	86	100.0	100.0	100.0	99.9	
Valid cases	82	Mean	3.07			

Q6 - LBB staff agree to agency requests to eliminate performance measures as needed.

	Frequency	Percent	Valid Percent	Greater than \$40M	Less than \$40M	Difference
Combined Agree	27	31.4	32.5	38.6	24.3	14.3
Neutral	26	30.2	31.3	22.7	40.5	-17.8
Combined Disagree	30	34.9	36.1	38.6	35.1	3.5
Missing	3	3.5	---	---	---	---
Total	86	100.0	100.0	99.9	99.9	
Valid cases	83	Mean	2.92			

Q7 - Performance measures change as agency functions change.

Value Label	Frequency	Percent	Valid Percent	Greater than \$40M	Less than \$40M	Difference
-------------	-----------	---------	---------------	--------------------	-----------------	------------

Combined Agree	41	47.7	50.0	54.8	42.1	12.7
Neutral	24	27.9	29.3	26.2	34.2	-8.0
Combined Disagree	17	19.8	20.7	19.0	23.7	-4.7
Missing	4	4.7	---	---	---	---
Total	86	100.0	100.0	100.0	100.0	
Valid cases	82	Mean	3.38			

Q8 - Performance data were used to modify measure definitions in the last two years.

	Frequency	Percent	Valid Percent	Greater than \$40M	Less than \$40M	Difference
Combined Agree	47	54.7	56.6	61.9	53.8	8.1
Neutral	14	16.3	16.9	19.0	12.8	6.2
Combined Disagree	22	25.6	26.5	19.0	33.3	-14.3
Missing	3	3.5	---	---	---	---
Total	86	100.0	100.0	99.9	99.9	
Valid cases	83	Mean	3.40			

Q9 - Performance data were used to modify measure targets in the last two years.

	Frequency	Percent	Valid Percent	Greater than \$40M	Less than \$40M	Difference
Combined Agree	65	75.6	76.5	82.2	71.8	10.4
Neutral	11	12.8	12.9	13.3	12.8	0.5
Combined Disagree	9	10.5	10.6	4.4	15.4	-11.0
Missing	1	1.2	---	---	---	---
Total	86	100.0	100.0	99.9	100.0	
Valid cases	85	Mean	3.89			

Q10 - Performance measures accurately represent important agency functions.

	Frequency	Percent	Valid Percent	Greater than \$40M	Less than \$40M	Difference
Combined Agree	43	50.0	50.6	53.3	48.7	4.6
Neutral	20	23.3	23.5	20.0	28.2	-8.2

Combined Disagree		22	25.6	25.9	26.7	23.1	3.6
Missing		1	1.2	---	---	---	---
	Total	86	100.0	100.0	100.0	100.0	
Valid cases	85	Mean	3.27				

Q11 - When all performance measures are considered together, they accurately represent overall agency functioning.

	Frequency	Percent	Valid Percent	Greater than \$40M	Less than \$40M	Difference
Combined Disagree	28	32.6	32.9	37.8	28.2	9.6
Neutral	21	24.4	24.7	22.2	25.6	-3.4
Combined Disagree	36	41.9	42.4	40.0	46.2	-6.2
Missing	1	1.2	---	---	---	---
	Total	86	100.0	100.0	100.0	
Valid cases	85	Mean	2.89			

Q12 - My agency has too many measures.

	Frequency	Percent	Valid Percent	Greater than \$40M	Less than \$40M	Difference
Combined Agree	33	38.4	38.8	46.7	28.2	18.5
Neutral	25	29.1	29.4	24.4	35.9	-11.5
Combined Disagree	27	31.4	31.8	28.9	35.9	-7.0
Missing	1	1.2	---	---	---	---
	Total	86	100.0	100.0	100.0	
Valid cases	85	Mean	3.19			

Q13 - My agency has too few measures.

	Frequency	Percent	Valid Percent	Greater than \$40M	Less than \$40M	Difference
Neutral	6	7.0	7.1	6.7	7.7	-1.0
Disagree	25	29.1	29.4	22.2	38.5	-16.3
Strongly Disagree	54	62.8	63.5	71.1	53.8	17.3
Missing	1	1.2	---	---	---	---
	Total	86	100.0	100.0	100.0	

Valid cases	85	Mean	2.28
-------------	----	------	------

Q14 - Staff at my agency have the necessary skills to develop quality performance measures.

	Frequency	Percent	Valid Percent	Greater than \$40M	Less than \$40M	Difference
Combined Agree	65	75.6	76.5	82.2	71.8	10.4
Neutral	10	11.6	11.8	6.7	17.9	-11.2
Combined Disagree	10	11.6	11.8	11.1	10.3	0.8
Missing	1	1.2	---	---	---	---
Total	86	100.0	100.0	100.0	100.0	

Valid cases 85 Mean 3.86

Q15 - Staff at my agency have the ability to interpret performance measure data.

	Frequency	Percent	Valid Percent	Greater than \$40M	Less than \$40M	Difference
Combined Agree	76	88.4	89.4	93.3	87.2	6.1
Neutral	6	7.0	7.1	4.4	7.7	-3.3
Combined Disagree	3	3.5	3.5	2.2	5.1	-2.9
Missing	1	1.2	---	---	---	---
Total	86	100.0	100.0	99.9	100.0	

Valid cases 85 Mean 4.11

Q16 - Staff at my agency have the ability to apply performance measure data.

	Frequency	Percent	Valid Percent	Greater than \$40M	Less than \$40M	Difference
Combined Agree	72	83.7	85.7	86.7	86.8	-0.1
Neutral	7	8.1	8.3	6.7	7.9	-1.2
Combined Disagree	5	5.8	6.0	6.7	5.3	1.4
Missing	2	2.3	---	---	---	---
Total	86	100.0	100.0	100.1	100.0	

Valid cases 84 Mean 4.00

Q17 - Performance data has been used for allocation of funds during the last two years.

	Frequency	Percent	Valid Percent	Greater than \$40M	Less than \$40M	Difference
Combined Agree	39	45.3	50.0	41.9	52.6	-10.7
Neutral	19	22.1	24.4	18.6	28.9	-10.3
Combined Disagree	20	23.3	25.6	34.9	13.2	21.7
Not Applicable	4	4.7	---	---	---	---
Missing	4	4.7	---	95.4	94.7	
Total	86	100.0	100.0			
Valid cases	78	Mean	3.32			

Q18 - Performance data has been used to improve agency services during the last two years.

	Frequency	Percent	Valid Percent	Greater than \$40M	Less than \$40M	Difference
Combined Agree	50	58.1	58.8	57.8	59.0	-1.2
Neutral	14	16.3	16.5	17.8	15.4	2.4
Combined Disagree	21	24.4	24.7	24.4	25.6	-1.2
Missing	1	1.2	---	---	---	---
Total	86	100.0	100.0	100.0	100.0	
Valid cases	85	Mean	3.39			

Q19 - Performance data has been used to improve contractor performance in the past two years.

	Frequency	Percent	Valid Percent	Greater than \$40M	Less than \$40M	Difference
Combined Agree	16	18.6	26.7	32.6	5.7	26.9
Neutral	27	31.4	45.0	34.9	31.4	3.5
Combined Disagree	17	19.8	28.3	20.9	22.9	-2.0
Not Applicable	19	22.1	---	---	---	---
Missing	7	8.1	---	88.4	60.0	
Total	86	100.0	100.0			
Valid cases	60	Mean	3.00			

Appendix E

APPENDIX E

Ratings and values legend

Sub-hypothesis Support Ratings and Values

No support – none of the survey items support sub-hypothesis. Value = 0.

Weak support – at least one, but fewer than half of the survey items support sub-hypothesis. Value = 1.

Majority Support – half or more, but not all survey items support sub-hypothesis. Value = 2.

Complete Support – all of the survey items support sub-hypothesis. Value = 3.

Hypothesis Support Ratings and Values

No support – average of associated sub-hypotheses values = 0.

Weak support – average of associated sub-hypotheses values is greater than 0 but less than 1.5.

Majority support – average of associated sub-hypotheses values is greater or equal to 1.5, but less than 3.

Complete support – average of associated sub-hypotheses values = 3.

Summary of Findings, values calculated – 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 avg. = 1
	WH1a	Broad input in performance measure development	majority support value = 2
	WH1b	Mutually responsive relationship between participants	no support value = 0
WH2	Sound scientific principles are used in performance measure use and development		mixed majority support avg. = 2.25
	WH2a	The process is ongoing, experimental and iterative	complete support value = 3
	WH2b	Measures are an accurate representation of agency function.	complete support value = 3
	WH2c	Arrays of measures accurately represent the overall mission	no support value = 0
	WH2d	Agencies have sufficient expertise in use of performance data.	complete support value = 3
WH3	Performance data are put to optimal use by state agency staff.		mixed majority support avg. = 2
	WH3a	Performance data are used to guide allocation of funds.	complete support value = 3
	WH3b	Performance data is used to improve agency services.	complete support value = 3
	WH3c	Performance are used to improve contractor performance.	no support value = 0

Summary of Findings, values calculated – 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 avg. = 2.25	
	WH4a	Community of inquiry principles are less evident in small state agencies compared to large agencies.		complete support avg. = 3	
		WH4a(1)	Broad input in performance measure development	complete support value = 3	
		WH4a(2)	Mutually responsive relationship between participants	complete support value = 3	
	WH4b	Scientific principles are less evident in small state agencies compared with large agencies			mixed majority support avg = 2.75
		WH4b(1)	The process is ongoing, experimental and iterative	complete support value = 3	
		WH4b(2)	Measures are an accurate representation of agency function.	complete support value = 3	
		WH4b(3)	Arrays of measures accurately represent the overall mission	complete support value = 3	
		WH4b(4)	Agencies have sufficient expertise in use of performance data.	majority support value = 2	
	WH4c	Use of performance data to guide agency operations is less evident in small state agencies compared with large agencies.			mixed weak support avg. = 1
		WH4c(1)	Performance data are used to guide allocation of funds.	no support value = 0	
		WH4c(2)	Performance data are used to improve agency services.	no support value = 0	
		WH4c(3)	Performance data are used to improve contractor performance.	complete support value = 3	