MEASURING COMPLIANCE PROGRAM PROGRESS AND IMPACTS: LESSONS FROM USEPA'S NATIONAL PETROLEUM REFINERY COMPLIANCE PROGRAM[1]

BUTLER, KATHLENE[1] AND FEKETE, GABRIELLE[2]

[1] 1200 Pennsylvania Avenue, Northwest, Mail Code 2460T, Washington, District of Columbia, 20009, USA, Butler.Katie@epa.gov

[2] 109 T.W. Alexander Drive, Research Triangle Park, North Carolina, 27709, USA, Fekete.Gabrielle@epa.gov

United States Environmental Protection Agency, Office of Inspector General, Office of Program Evaluation, Cross-Media Issues

SUMMARY

An evaluation of the US Environmental Protection Agency's (USEPA) National Petroleum Refinery Program demonstrates how choosing performance measures that do not measure program impacts can lead to uncertainty about results. Public reports on the petroleum refinery program claimed that the program eliminated 200,000 tons of pollutants from the air surrounding refineries every year. However, USEPA and refineries reported predicted emissions reductions instead of actual reductions achieved. Moreover, USEPA did not actually know how much pollution participating companies had eliminated. USEPA claimed the program a success based on the modeling of predicted emissions reductions, most of which would not occur for 10 years or more, and the number of companies participating in the program. This case study demonstrated the challenge program managers face in choosing performance measures that clearly connect outputs, intermediate outcomes, and end outcomes. Enforcement personnel must demonstrate that their program improved compliance within the targeted population, or that the risk posed by the population decreased. Given the challenge of choosing measures that capture the data most indicative of a program's relative success or failure, the USEPA Office of Inspector General (OIG) developed a list of key assessment questions for performance measures so that program managers can improve their performance measurement schemes. Thereafter, the OIG recommends for testing the key assessment questions by evaluating sample enforcement programs in the spring of 2005.

1 INTRODUCTION

In this paper, we discuss the importance of performance measurement for environmental enforcement and compliance programs and outline criteria for developing and assessing performance measures. We discuss how performance measurement for the USEPA National Petroleum Refinery Compliance Program could be improved to better demonstrate program outcomes (based primarily on the 22 June 2004 USEPA Office of the Inspector General evaluation report EPA Needs to Improve Tracking of National Petroleum Refinery Compliance Program Progress and Impacts, Report No. 2004-P-00021). Finally, we describe a set of key questions

for assessing existing performance measures and performance measurement suites developed for environmental enforcement and compliance programs, and make recommendations for testing the questions and subsequent use.

1.1 Challenge of Performance Measurement for Environmental Enforcement and Compliance Activities

Public officials responsible for environmental enforcement and compliance personnel face unique performance measurement challenges: (1) they must demonstrate that their programs improve compliance with regulations across the entire regulated community in order to demonstrate that their programs improve regulatory compliance, in general; (2) at the same time, they must demonstrate that they target significant non-compliers to ensure that their activities address the most significant risks to human health and the environment.[2] They must meet these challenges and report results in a way that enhances their credibility and accountability with lawmakers, the public, and the regulated community.[3]

With over 40 million regulated entities in the United States,[4] demonstrating improvements in compliance across all regulated entities is a challenging endeavor for the USEPA. USPEA allocates scarce monetary and staff resources between randomized inspections (for determining statistically-valid compliance rates) and targeted inspections (for ensuring that they regularly inspect the most high-risk violators). In addition, USEPA must measure these conditions in the face of a multitude of additional factors contributing to environmental compliance – prevailing economic conditions and market forces, inconsistent and competing local, state, and federal priorities, environmental advocacy and citizen groups' efforts, and media attention among others.[5]

Recently, some public watchdog groups have criticized USEPA for a reported decline in enforcement actions, saying that the agency relaxed its enforcement efforts.[6] However, others could have attributed the reported decline to increased compliance – perhaps fewer enforcement actions were necessary because fewer regulated entities required enforcement. Counting how many people or companies the agency found in violation gives no information about the severity of the violation, whether the facility corrected the violation, whether the company was a repeat-offender, or whether Americans and their natural environments will be safer because of the enforcement action.[7]

Because USEPA did not have measurable environmental results to complement its enforcement and compliance assurance claims, it faced pressure to increase the number of enforcement actions.[8] USEPA officials recognized that they should improve measurements of enforcement and compliance accomplishments, and did not want to be held to preconceived expectations that more enforcement actions meant better compliance in the regulated community.[9]

Experts in performance measurement and regulatory programs also stress the importance of effective performance measures. In operating integrated compliance programs, Kiener, et al. said that effective planning depended on agencies' abilities to gather data, measure performance, and monitor environmental conditions.[10] They said doing these things would enable an agency to establish baselines, identify and prioritize compliance problems, and manage programs in response to incoming performance information. Metzenbaum agreed, saying that a good performance measurement system holds an organization accountable, and improves outcomes, including increasing awareness, sharpening focus, motivating improved performance, encouraging innovation, and allowing for adaptation in response to results.[11]

Behn says performance measures enable public managers to accomplish several necessary tasks: evaluation, control, budgeting, motivating staff and stakeholders, promoting their programs, celebrating

successes, learning about what works and what does not work, and improving upon programs based on that information.[12] In addition, Sparrow said regulatory enforcement performance measurement systems cover six factors: (1) customer or client satisfaction, (2) employee satisfaction, (3) likelihood of identifying a complier as a non-complier or not correctly identifying a non-complier, (4) ensuring that most activities are aimed at high-risk groups and noncompliance is meaningful (contributes to environmental or human health problems), (5) measuring internal productivity, and (6) measuring efficiency.[13]

To ensure that they adequately characterize program performance, program managers can use logic modeling or a similar tool to demonstrate logical connections between outputs (also called activity counts, such as the number of facilities inspected), intermediate outcomes (such as pounds of pollution reduced by an enforcement action), and end outcomes (human health improvements as a result of an enforcement action).

By using these and other criteria to choose, change, and use performance measures, enforcement and compliance program operators can manage programs based on results to ensure they use the best techniques and achieve the best possible outcomes.[14]

2 DEVELOPMENT OF USEPA'S PETROLEUM REFINERY STRATEGY

USEPA began a targeted strategy in the petroleum refinery sector in 1996 because they accounted for significant releases of pollution into the environment. In 2001, refineries released over 35,000 tons of toxic air pollutants according to USEPA's Sector Facility Indexing Project data (a publicly available on-line database retired in 2004). In 1999, according to the most current data from USEPA's AirData system, refineries released approximately 243,000 tons of nitrous oxides, 396,000 tons of sulfur dioxide, and 412,000 tons of other common air pollutants.[15]

Petroleum refinery emissions seriously impact human health and the environment. In 2000, USEPA reported that 45 percent of all refineries at that time were within 3 miles of population centers containing 25,000 or more people, and 26 percent were within 3 miles of population centers containing 50,000 or more people. Varying environmental and human health effects resulted from the following common air pollutants released at refineries: volatile organic compounds, sulfur dioxide, nitrous oxides, particulate matter, carbon monoxide, hydrogen sulfide, and toxic air pollutants. Toxic air pollutants include pollutants known or suspected to cause cancer or other serious human health effects.[16]

USEPA and the U.S. Department of Justice developed and implemented a petroleum refinery compliance strategy to address important noncompliance problems in the industry. USEPA and regional officials used inspections, formal USEPA information requests to refineries, and industry trade journals to identify refinery priority areas. USEPA used the results of these initial research efforts to focus (or target) investigations on the noncompliance areas indicated by their research. USEPA's national experts continued gaining experience regarding compliance issues within the refinery industry, and helped select the four Clean Air Act priority areas that became the refinery program's focal point.[17]

USEPA's national refinery compliance program evolved as USEPA learned more about the noncompliance issues and applied various tools and strategies to address those issues. In 2000, USEPA began pursuing voluntary global settlements with refinery companies that resulted in consent decrees. USEPA and regional officials coordinated with the U.S. Department of Justice, who led all of the global settlement negotiations. USEPA offered corporate officials the opportunity to avoid possible investigation and litigation by signing consent decrees. USEPA's strategy included coordinating with interested States and local authorities.

The first two consent decrees entered the implementation phase in early

2001. The consent decrees spanned 8 to 10 years and required coordination and communication among USEPA, USEPA regions, States, and industry. The signing of a consent decree ended the settlement process for that company and began a new process of oversight by USEPA and interaction between USEPA, states, and the companies. As of 2004, USEPA continued to conduct negotiations, assist regions in assuming a larger role with the refinery program, and work with refiners to implement consent decrees.[18]

3 METHODOLOGY AND SCOPE

We evaluated USEPA's national petroleum refinery program between June 2003 and March 2004 to determine what impact the program had on compliance among refineries.

To understand the nature and extent of the petroleum refinery universe and what strategies USEPA and its partners developed to address compliance at refineries, we interviewed and collected documents from USEPA staff, USEPA's National Enforcement Investigations Center, regions, States, industry, environmental groups, and the U.S. Department of Justice. To determine what impact the program had on compliance among refineries, we evaluated the performance measurement and reporting approach for petroleum refineries by interviewing staff in USEPA's Office of Regulatory Enforcement, Office of Compliance, and USEPA refinery issue experts in Headquarters, USEPA Region 5, and the National Enforcement Investigations Center, and by analyzing supporting documentation.[19]

After discovering inadequacies in refinery program performance measures, we determined to develop key questions for environmental enforcement and compliance performance measure assessment. In developing the key questions, we reviewed public policy literature to summarize up-to-date criteria for developing and assessing performance measures (see References). We also interviewed national performance measurement, environmental, and enforcement and compliance experts to discuss assessment tools' criteria, and to inquire about any additional criteria that may be appropriate. These included individuals from government agencies, nonprofit policy groups, and universities. Our literature summary and interviews provided the basis for developing the key assessment questions for the identification of strengths and weaknesses of measures.

4 RESULTS

USEPA's strategy to improve compliance at US petroleum refineries reported outputs and predicted intermediate outcomes, but did not track or report actual results (intermediate or end outcomes). The outputs and predicted intermediate outcomes demonstrated success in reaching settlements and getting companies to promise emissions reductions. However, because USEPA did not measure or report outcomes, it did not have information about the actual program results. In fact, a recent investigative news report indicated that refineries in the program were not achieving the promised emissions reductions, and that EPA was not reporting that information to the public.[20] If program managers assessed the refinery program's performance measures during the program, USEPA could better determine what program improvements would increase accountability and credibility for the program among regulated entities, lawmakers, and the public. They would have been able to demonstrate whether their program improved compliance within the targeted population, or whether the risk posed by the population decreased as a result of the program.[21]

To provide sound criteria for assessing performance measures, we summarized criteria and suggested key questions for assessing performance measures so that program managers can ensure that they are measuring the best indicators of success.

4.1 Settlements Projected to Result in Significant Emissions Reductions[22]

By January 2005, USEPA had entered into 12 global settlements (or consent decrees) covering 48 of the 145 refineries. The settlements covered more than 40 percent of total U.S. petroleum refining capacity. Based on settling companies' estimates, EPA predicted it would achieve annual atmospheric emissions reductions of approximately 50,000 tons of nitrous oxides and 120,000 tons of sulfur dioxide, as well as reductions in benzene, volatile organic compounds, and particulate matter once companies fully implemented the consent decrees. The settling companies agreed to invest more than $2.1 billion in pollution control technologies and pay civil penalties of $40.4 million. These refineries also agreed to implement supplemental environmental projects valued at approximately $30 million. In exchange, USEPA offered a "covenant not-to-sue," or a release from liability for any pre-consent decree regulatory violations associated with the four priority areas.

The global settlements also relieved USEPA from having to conduct resource-intensive investigations at each refinery a company owned. According to USEPA, a refinery-by-refinery, issue-by-issue approach, in which USEPA conducted an individual inspection or investigation at each and every refinery followed by information requests, notices of violation, negotiations and/or litigation, could take many years and require resources beyond USEPA's means. The refinery consent decrees required each company to take various actions over the next several years. These actions include implementing air pollution controls as well as developing policies and procedures that go beyond compliance with existing regulations. In addition, USEPA and the companies agreed to test and use innovative technologies.

4.2 Performance Measures Focused on Outputs and Projected Outcomes[23]

The performance measures tracked internally and reported to the press focused on outputs, such as the number of companies in consent decrees and the percent of the refining capacity covered by consent decrees, and on projected rather than actual environmental outcomes. USEPA reported program results in two ways:

First, USEPA reported results to the public through press releases. USEPA used press releases to communicate the signing of consent decrees to the public, the projected emissions reductions at full implementation of consent decrees (consent decrees lasted 8 to 10 years), and the dollars companies agreed to pay in penalties as a result of consent decrees. USEPA management did not plan to issue press releases or other reports to the public detailing the end outcomes of consent decree implementation because USEPA management did not believe the press would be interested.

Second, USEPA reported results to Congress using a compliance information system. For refinery consent decrees, USEPA input data into the system representing (1) the projected annual emissions reductions that would be realized once implementation was complete, (2) the dollar amount of penalties generated, and (3) the dollar value of required supplemental environmental projects. According to USEPA, the system was not designed to capture, and did not capture, information about environmental outcomes from the consent decrees, such as demonstrated environmental and human health benefits.

USEPA used three systems for collecting information on consent decree implementation for internal use: (1) company data collected through consent decree reports, (2) monthly conference calls between USEPA managers and staff working on consent decree implementation, and (3) a contractor-developed consent decree tracking system. However, USEPA did not use these systems to demonstrate progress toward meeting consent decree goals.

Consent decrees required compa-

nies to provide quarterly reports that included actual emissions data related to the most significant emissions issues contained in consent decrees. USEPA used the information to set some emissions limits that consent decrees did not specify, but did not use this information to monitor, verify, or report progress toward achieving consent decree goals. An assessment of their performance measures could have led USEPA to incorporate this existing information into their performance measurement system for the refinery consent decrees, providing them with actual knowledge about the state of emission reductions and the ability to make program changes as necessary to improve compliance.

USEPA's internal and external performance measurement system did not account for intermediate outcomes or end outcomes, even though the effort relied on new, unproven technologies, and even though USEPA reported predictions to the public.

4.3 External Report Suggested Refineries Were Not Achieving Projected Results

A 12 December 2004 investigative report in the Fort Worth [Texas] Star Telegram indicated that USEPA allowed two of every three companies to miss consent decree deadlines, and did not notify the public, courts, or local pollution control authorities, as required by the settlements. Therefore, the Star Telegram said, USEPA had not achieved promised environmental benefits. They found that the program had reduced about 40,000 tons of nitrous oxides, sulfur dioxide, and particulate matter combined from 2001 to 2004, while USEPA claimed in the press that reductions totaled 200,000 tons per year. When pressed, USEPA said it did not actually know how much pollution the initiative had reduced.

Further, the Star Telegram found that USEPA was not sure that some technologies would work when companies signed the consent decrees, and some technologies were not reducing pollution as predicted. Although several companies told USEPA that a new technology employed did not have the intended pollution-reducing effects, USEPA continued to require the technology in subsequent consent decrees. The article said that the refinery program gave "'the illusion of progress without actual progress'."[24]

4.4 A Method for Assessing Enforcement and Compliance Performance Measures

To ensure that compliance programs achieve the intended results, program managers should be able to periodically assess their performance measures to ensure that they are still reliable, relevant, feasible, and comparable with others' efforts. If program managers in the refinery program assessed the program's performance measures, they could have made improvements to consent decree outcomes and could have improved subsequent negotiations based on accessible outcome information reported to the agency.

Experts considered USEPA's enforcement and compliance measurement program to be a national and international model.[25] However, while USEPA assessed its performance measures for enforcement and compliance based on internal criteria, neither USEPA nor the literature had a systematic methodology for assessing enforcement and compliance performance measures. We subsequently developed a series of key assessment questions based on the literature, and a scoring system for determining strengths and weaknesses based on key criteria (Appendix A).

4.4.1 Key Performance Measure Assessment Questions

Below, we summarize major issues in performance measure adequacy and propose questions for program mangers that address these issues.

4.4.1.1 Measuring Outcomes

Performance measurement and environmental professionals agree that

performance measures should relate to intended outcomes and activities as closely as possible. They say that users should be able to clearly understand performance reports that organizations present to them. Existing literature recommends developing performance measures that are relevant to overall program goals, policies, management decision-making, other users' needs (both user-friendly and user-focused), and intended results, and that provide connections between activities and outcomes.[26] However, the connection between program activities and end outcomes is especially difficult to prove in enforcement and compliance programs (due to other influencing factors and the sometimes-substantial time lag between program activities and environmental results), so program managers should focus on demonstrating their success at focusing programs on risks and subsequently decreasing the risks identified.[27] By describing how they connect activities with inputs, outputs, intermediate outcomes, and end outcomes, program managers can demonstrate how different types of measures establish connections between activities and outcomes, including organizational goals and objectives.[28] To address connections to outcomes, we ask the following questions:

—Are the measure's connections to organizational goals and objectives demonstrated through logic modeling or a similar method?

—Is the measure categorized as an input, output, intermediate outcome, or end outcome measure?

—Are the important aspects of performance (based on program goals) included in the family of measures (such as change in compliance for the targeted population, or change in risk posed by the targeted pollutant)?

—Do measures meet needs of users (such as geographically-specific information)?

4.4.1.2 Making Measures Useful

In order for performance to actually improve through the use of performance measures, program operators must give constant attention to measures – as frequently as on a weekly basis,[29] using the information to improve the program as necessary. In addition, public use of performance measures requires that managers, lawmakers, and the public have access to information within a reasonable time period. To address data use, we ask the following questions:

—Does the suite of measures provide all the information that is important for program decision-making?

—Are compliance data reported within a timeframe that allows users to take action or make decisions based on the results?

—Are stakeholders' comments on compliance data, measures, and reporting systematically considered (e.g., collected, assessed, and addressed)?

—Are there plans to incorporate, develop, or implement compliance measures into the suite that would improve relevance/reliability of the suite?

4.4.1.3 Using Accurate Data

To make sure that program managers make decisions based on analytically-sound information, we ask the following questions:

—Did the activities and results reported actually occur in the time period indicated?[30]

—If the measure is based on modeling or predictions, has it been verified to ensure reliability and validity of the estimate?

—Was counting of transactions, conditions, and events accurate and complete (no over- or under-counting)?[31]

—Did the measure adequately represent the population so that inferences can be drawn?[32]

—Were performance measure data calculations correct (e.g., computations

of field data, incorporations into measure reports, etc.)? (Assessor should check a sample of data for accurate computation from collection sheets through incorporation into measure reporting.)[33]

—Are data from different sources (e.g., offices, projects, or organizations) based on similar definitions of compliance and compliance assessment procedures, or did you make them comparable?[34]

—Do repeated compliance measurements or measurements by different parties (e.g., states versus EPA) yield the same results?[35]

—Has the organization recognized and addressed known flaws or errors in collecting, reporting, presenting, or transcribing data that is used in performance measures?[36]

—Are there systems in place to detect abnormalities in compliance measure values?[37]

—Are there adequate procedures in place that ensure compliance data records are not altered, lost, or incorrectly transferred (e.g., data storage, certification)?[38]

4.4.1.4 Making Measures Feasible

Although recent efforts to assess government programs in the United States focused heavily on measuring ultimate outcomes, like "human lives saved",[39] this may not be possible to measure using current data sources and technologies, or given current funding or training limitations. Agencies could consider including this type of measure as "developmental" to indicate that they intend to move toward measuring it. To address this issue, we ask the following questions:

—Are data required for the measure measurable (ie. does the technology exist for measuring)?

—Are the costs associated with the measure achieving maximum benefits?

—Is there sufficient staff and funding for data collection, analysis, and reporting?

—Do staff collecting the information have the training to accurately measure compliance —e.g., collect sufficiently complete, consistent, and accurate data?

4.4.1.5 Making Measures Comparable

To compare results with those from other, similar programs, measures should be comparable with those used by other major colleagues' programs. For example, the North American Agreement on Environmental Cooperation has developed cooperative enforcement programs across Canada, the United States, and Mexico. Communication of compliance information using similar measures or indicators requires that the countries involved generally agree about what should be measured.[40] Differing definitions of "compliance" among reviewers could lead to inconsistent inspections; for example, non-compliers may accidentally be identified as "in compliance", or compliers may accidentally be identified as non-compliers.[41] If this happens, decisions about enforcement and compliance activities may be made using inaccurate compliance information. To address comparability, we ask the following questions:

—Can compliance measures be compared against baselines, previous performance, or others who conduct similar activities?

—Are data from different states, countries, or other pertinent organizations based on similar definitions of compliance and compliance data collection procedures?

—To address this issue, we asked the following question:

—Are data from different offices, projects, or organizations based on similar definitions of data elements and data collection procedures?

4.4.1.6 Reporting Performance Measures

In preparing external reports on performance measures, program man-

agers should provide enough information for users to correctly understand results, including information about how present performance compares with past performance or performance against a baseline, and explanations of results.[42] To address reporting, we ask the following questions:

- Are performance measures presented along with comparison data that illustrate the adequacy of performance (e.g., performance trends, performance against benchmarks)?
- Are changes in performance, methodology, or compliance explained?

Are data limitations (e.g., missing data, lag time, etc.) described in performance communications?

- Is the measure explained, interpreted, and presented, so users can understand what the measurements say about changes in compliance?

By asking these key questions, program managers will have considered the key criteria for ensuring that performance measures are feasible, and provide reliable, relevant, and comparable information about on-going programs.

4.4.2 Proposed Method for Determining Strengths and Weaknesses

After assessing how well measures meet the criteria listed in the previous section, we propose that managers score answers based on five overall criteria: relevance, accuracy, feasibility, comparability, and reporting. Appendix B offers a template for scoring according to these criteria, based on the answers chosen on the questionnaire presented in Appendix A. We determined that neither questions nor overall criteria would be weighted during scoring to enable program managers to assess the relative importance of each question and criterion for their programs as they see fit. The end result for a measure scored using the proposed system would be a series of histograms for each category. Program managers could assess how well a given measure met each criterion represented. Finally, once program managers assessed a suite of measures for a given program, they could look across the histograms for each criterion to determine which measures met which criteria. This information would help them to make specific decisions about how to improve existing measures or add additional measures to better demonstrate results, and to improve credibility and accountability for the program.

5 CONCLUSIONS

Successful enforcement and compliance programs change the performance of targeted companies, including behaviors (such as implementing environmental management systems) as well as actual compliance with regulations.

Improved performance measures for the USEPA National Petroleum Refinery Compliance Program could have better demonstrated changes that resulted from the program (outcomes), and alleviated perceived problems in public accountability posed by reporting predicted outcomes. Consent decrees required companies to provide quarterly reports that included actual emissions data related to the most significant emissions issues contained in consent decrees. USEPA used the information to set some emissions limits that consent decrees did not specify, but did not use this information to monitor, verify, or report progress toward achieving consent decree goals. An assessment of program performance measures could have led USEPA to incorporate this existing information into their performance measurement system for the refinery consent decrees, providing them with actual knowledge about the state of emission reductions and the ability to make program changes as necessary to improve compliance.

In order to ensure that program managers choose the best measures of program performance, they should assess performance measures on a regular basis to determine how well they meet established criteria. The key assessment ques-

tions and scoring system we proposed based on extensive performance measurement and regulatory program literature could aid program managers in such an assessment. The USEPA Office of Inspector General intends to test the key assessment questions and scoring system by evaluating a sample of enforcement and compliance measures at USEPA. We intend to validate our assessment by asking a panel of enforcement and compliance personnel to duplicate our evaluation and offer comments on ease of use and perceived accuracy for the assessment tool.

6 REFERENCES

[1] This work would not have been possible without the generous contributions of Erin Barnes-Weaver, Kim Bryant, Andrew Creath, Jeffrey Harris, PhD, Director, Jeff Hart, Assignment Manager, and Benjamin Webster, PhD, all of the United States Environmental Protection Agency, Office of Inspector General, Office of Program Evaluation, Cross-Media Issues.

[2] Sparrow, M. K., The Regulatory Craft: Controlling Risks, Solving Problems, and Managing Compliance. Council for Excellence in Government, Brookings Institution Press: Washington, DC, 2000.

[3] Metzenbaum, S. 1998. Making Measurement Matter: The Challenge and Promise of Building a Performance-Focused Environmental Protection System. The Brookings Institution.

[4] www.epa.gov/oeca/

[5] Hatry, H. 2004. Personal communication.

[6] See, for example: Environmental Integrity Project. States Forced to Act as Bush Administration Fails to Control Air Pollution. January 12, 2004.

[7] Sparrow, M. K. 2000.

[8] Sparrow, M. K. 2000.

[9] Metzenbaum, S. 2003. More Nutritious Beans. In The Environmental Forum. The Environmental Law Institute: Washington, DC; and Stahl, M. 2003. Performance Indicators for Environmental Compliance and Enforcement Programs: The U.S. EPA Experience. In: Measuring What Matters, Proceedings from the INECE_OECD Workshop on Environmental Compliance and Enforcement Indicators, OECD Headquarters, Paris, France. 3-4 November 2003;

[10] Kiener, S., L. Paddock, and M. Stoughton. 2003. Beyond Enforcement? Environment, Compliance Assistance, and Corporate Leadership Programs in Five Midwest States. Tells Institute, Environmental Law Institute and National Academy of Public Administration.

[11] Metzenbaum, S. 1998.

[12] Behn, R. D. 2003. Why Measure Performance? Different Purposes Require Different Measures, Public Administration Review, 63(5): 586-606.

[14] Sparrow, M. K. 2000.

[14] Stahl, M. M. 2003.

[15] USEPA Office of Inspector General. EPA Needs to Improve Tracking of National Petroleum Refinery Compliance Program Progress and Impacts, Report No. 2004-P-00021. 22 June 2004

[16] Ibid.

[17] Ibid.

[18] Ibid.

[19] More information about methodology and scope for the OIG evaluation of the USEPA refinery program can be found in USEPA Office of Inspector General. EPA Needs to Improve Tracking of National Petroleum Refinery Compliance Program Progress and Impacts, Report No. 2004-P-00021. 22 June 2004.

[20] Streater, S. EPA Hasn't Won the Improvements it Touted, Ft. Worth Star Telegram, December 12, 2004.

[21] Sparrow, M. K. 2000.

[22] Section 3.1 from USEPA OIG 2004.

[23] Section 3.2 from USEPA OIG 2004.

[24] Streater, S. 2004.

[25] Sparrow, personal communication, January 10, 2005; Markowitz, Ken, INECE, personal communication, August 31, 2004.

[26] Linster, M. 2003. Environmental Indicators: Development, Measurement and Use. In: Measuring What Matters, Proceedings from the INECE_OECD Workshop on Environmental Compliance and Enforcement Indicators, OECD Headquarters, Paris, France. 3-4 November 2003; Discussion Paper: INECE-OECD Workshop on Environmental Compliance and Enforcement Indicators: Measuring What Matters; Soniat, E. 2004. Assessing the Reliability and Relevance of Performance Measures, USDA Graduate School Course AUDT9201; Keel, J. and Hawkins, A., 2000, State of Texas Guide to Performance Measure Management, 2000 Edition; Wholey, J. 1999. Quality Control: Assessing the Accuracy and Usefulness of Performance Measurement Systems. In Performance Measurement: Getting Results by Hatry, H., The Urban Institute.; Metzenbaum, S. 1998; Keiner, S., et al., 2003.; Morgan, S. and Slusher, D. 2002. Appendix D: Performance Measurement and Reliability Issues, Memo from Austin City Auditor, and, City of Austin Managing for Results Business Planning Guide (revised 2003). Austin City Budget Office; Mihm, J. C. 2000. GAO Letter Report B-285312 Subject: Managing for Results: Assessing the Quality of Program Performance Data. Chemical and Pesticides Results Measures project. CAPRM II Project Document: Chemical and Pesticide Result Measures II. Accessed on December 7, 2004 at http://www.pepps.fsu.edu/CAPRM/doc2.html; Florida Department of Environmental Protection, Office of Inspector General. 1998. Performance Based Program Budgeting Assessment Blueprint. Report No. IA-06-01-98-002; and, Stahl, M. 1997. Measuring the Performance of EPA's Enforcement and Compliance Assurance Program: Final Report of the National Performance Measures Strategy. USEPA.

[27] Sparrow, 2000.

[28] Sparrow, 2000.

[29] Behn, R. D. On the ludicrous search for: The Magical Performance System. Bob Behn's Public Management Report. Vol. 2 no.5 (January 2005).

[30] Soniat, 2004.

[31] Ibid.

[32] Ibid.

[33] Soniat, 2004, and Florida Department of Environmental Protection 2002.

[34] Soniat, 2004.

[35] Ibid.

[36] Florida Department of Environmental Protection 2002.

[37] Ibid.

[38] Ibid.

[39] United States Office of Management and Budget, Program Assessment Rating Tool (PART) available at http://www.whitehouse.gov/omb/part/index.html and http://www.whitehouse.gov/omb/part/2004_program_eval.pdf.

[40] Harman, S. A. and L. I. Lawrence. Emerging Networks of Environmental Enforcement and Compliance Cooperation in North America and the Western Hemisphere. In: Measuring What Matters, Proceedings from the INECE_OECD Workshop on Environmental Compliance and Enforcement Indicators, OECD Headquarters, Paris, France. 3-4 November 2003.

[41] United States General Accounting Office. 1998. Head Start. Challenges in Monitoring Program Quality and Demonstrating Results. Report to Congressional Requesters, June 1998; Linster, M. 2003; Sparrow, 2000.

[42] Tate, R. L. 2004. Auditing Performance Measures: Lessons Learned, AGA Conference; and Tate, R. L. 2004. Managing for Results: Performance Measure Certification, Maricopa County, Arizona, Audit Report.