Spring 1-1-2013

A Methodology for Building Evaluation Capacity in Alternative Fuel Deployment Programs

Nathaniel John Sobin
University of Colorado at Boulder, sobin@colorado.edu

A METHODOLOGY FOR BUILDING EVALUATION CAPACITY IN ALTERNATIVE FUEL DEPLOYMENT PROGRAMS

by

NATHANIEL JOHN SOBIN

B.S., University of Arkansas at Fayetteville, 2004

M.S., University of Arkansas at Fayetteville, 2005

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirement for the degree of

Doctor of Philosophy

Department of Civil, Environmental, and Architectural Engineering

2013

This thesis entitled:
A Methodology for Building Evaluation Capacity in Alternative Fuel Deployment Programs
written by Nathaniel John Sobin
has been approved for the Department of Civil, Environmental, and Architectural Engineering

____
(Keith Robert Molenaar, Committee Chair)

____
(Matthew Ryan Hallowell, Committee Member)

Date:____________

The final copy of this thesis has been examined by the signatories, and we
Find that both the content and the form meet acceptable presentation standards
Of scholarly work in the above mentioned discipline.

IRB protocol # _12-0740_

Sobin, Nathaniel (Ph.D., Department of Civil, Environmental, and Architectural Engineering)

A Methodology for Building Evaluation Capacity in Alternative Fuel Deployment Programs

Thesis directed by Professor Keith Robert Molenaar

Many agencies have and continue to deploy alternative fuel (AF) technologies for the on-road transportation sector with the goal of addressing multiple public objectives including reduced environmental impact, increased energy security, stimulating economic growth, and stimulating technology transition. However, there is little evidence of agreement by agencies on which objective, or combination of objectives, should be addressed or how to best prioritize goals. Furthermore, there is little agreement on how to reliably and consistently measure progress toward desired outcomes once goals are defined. These issues are contributing factors to the current lack of deployment where alternative fuel vehicles (AFVs) represent only a small fraction of vehicles on the road in 2013. This research provides a means for addressing these issues through the broader concepts of evaluation capacity building (ECB). In this research, ECB refers to the level of discontinuity in the prioritization of goals (goal ambiguity) and the level of discontinuity and reliability within metrics used to gauge progress towards those goals. These concepts are demonstrated through an analysis of evaluation criteria used in fuel- and project-neutral (wide-scope) grant programs. This research shows that when agencies deploy alternative fuel technologies, environmental goals are most commonly targeted and are similarly evaluated among AF deployment grant programs at federal, state, and regional levels. In contrast, varying levels of goal ambiguity exist in the goal domains of energy security, economic growth, and technology transition. This research also demonstrates that there are significant differences in the performance of evaluation criteria currently in use. The results show that units

of fuel displacement, the number of alternatively fueled vehicles distributed, and the number of alternative fuel refueling stations installed are the most reliable metrics currently incorporated within deployment programs. Conversely, other metrics commonly in use today present challenges in terms of the feasibility of acquiring the necessary data, the objectivity with which data are reported, and the clarity of definitions used for these metrics. Collectively this research provides a framework for consistently and holistically measuring evaluation capacity between alternative fuel deployment programs.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

OBSERVED PROBLEM

The advancement of alternative fuels (AFs) by governmental programs is a common method used to address multiple objectives including decreased environmental impact, accelerated technology transition, spurred economic growth, and enhanced energy security. Yet as various levels of government in the United States devise programs to advance AF technologies to address any one, or combination of these goals, there is little consensus on how to prioritize these goals or how to reliably measure progress towards their achievement. Despite considerable and sustained deployment efforts over the past 40 years, alternative fuel vehicles (AFVs) represent less than two percent of light-, medium-, and heavy-duty vehicles on the road today (USEIA, 2013). This is, at least in part, a consequence of AF program designs that overlook the benefits of a strategic approach between deployment programs and between projects as well as the development of valid metrics to measure progress.

The lack of a strategic approach in alternative fuel deployment is a concept well documented in literature. Literature routinely articulates the need for such an approach through an outcry for consistent, aligned, and prioritized messages, as well as consistent evaluation of performance, from governments at state, regional, federal, and even international levels (USGAO, 2011; Deutch, 2011; IEA, 2011; NREL, 2011; Sperling and Gordon, 2009; Sperling and Yeh, 2009; Lee, 2009; Melaina et al., 2008; Melendez, 2006; Sagar and Gallagher, 2004; Byrne and Polonsky, 2001; IEA, 2002; Howell and Chelius,1997; Sperling, 1988). While there is a well-defined need for such an approach, there is little consensus on how such an approach is designed or implemented. Within the need for a strategic approach lie several underlying concepts including the alignment of goals, the alignment of evaluation systems, and the use of reliable and valid metrics within those evaluation systems. Each of these concepts necessary to create the greater strategic approach has a unique set of problems and potential benefits.

The alignment of goals is part of the larger concept of goal setting theory as described and researched by Locke, where higher and more clearly defined goals result in higher performance in terms of human motivation (Locke, 1968). Goal alignment is a commonly sought objective within nearly all organizations and systems of programs, and with good reason. Literature often implies the potential benefits that goal alignment can provide, including both increased efficiency within an organization and increased output from an organization (Burns and Stalker, 1961; Woodward 1965; Lawrence and Lorsch, 1967; Miles, et al., 1978; Mintzberg, 1979; Ouchi, 1980; Miller and Friesen, 1984; Witt, 1998; Higgins and Maciariello, 2004). As a result goal alignment continues to be an objective sought by organizations and remains a common research interest.

Alternative fuels for the on-road transportation sector are used to address many goals including decreased environmental impact, increased energy security, increased economic growth, and an accelerated transition towards AF technologies (Deutch, 2011; Sperling and Gordon, 2009; IEA 2002; Sperling, 1988). However, policies and programs designed to deploy AF technologies are commonly prone to misalignment simply by attempting to address multiple goals and objectives (Deutch, 2011; Sperling and Gordon, 2009; Sperling, 1988). Deutch emphasizes the presence and subsequent complication of multiple goals inherent to alternative energy policy when he states "*Clarity about the purpose of energy policy is also important. It is easy to have a single goal and complicated to have multiple goals, especially when the combination is intended to overwhelm any doubt about the virtue of the policy*" (Deutch, 2011, p.15). However, the realities of policy and program development ring true when he states that "…*a single objective—for* *example, reducing emissions—is simplest, but multiple objectives are the rule…*" (Deutch, 2011, p.16). In addition, there is an extremely diverse group of stakeholders involved in AF deployment, with multiple needs and objectives and with differing means for expressing them (Byrne and Polonsky, 2001; Sperling and Gordon, 2009). As a result, AF policies end up incorporating a patchwork of intentions and goals expressed in a variety of goals and policies that are occasionally divergent, if not entirely conflicting. Despite these difficulties, and the subsequent deleterious consequences, there is little research that describes goal alignment within AF deployment programs.

The evaluation of performance towards goals is also a major portion of Locke's goal-setting theory (Locke, 1968). In performance evaluation literature, consistent and reliable evaluation methods and metrics are shown to again lead to higher inter- and intra- organization performance (see Saad, 2001 for a detailed discussion on this topic). As a result, evaluation alignment is also an often sought portion of strategic management within nearly every kind of organization. For better or worse, policy and programmatic evaluation has become the "norm" for nearly all governmental programs, including AF deployment programs (Rainey, 2009; Moynihan, 2008; USDOE, 2011). Such requirements are largely a product of legislation such as the Government Performance and Results Act (GPRA) of 1993 and are executed through systems such as the Performance Assessment Rating Tool (PART) that is overseen by the Office of Management and Budget (OMB) (USC, 1993; Moynihan, 2008; Rainey, 2009). As a result, consistent and reliable evaluation of policies and programs has become a major part of nearly all governmental programs, including AF deployment programs, from the standpoint of both improving performance and mandate.

Within AF deployment programs, there are fragmented and diverse methods for evaluating performance towards goals. A study performed by the International Energy Agency (IEA) on alternative fuel deployment policies and programs states that, *"… the fact that even the National experts of the Annex were not able to identify programme evaluations or to report their results, is very telling regarding the status of and utilisation of evaluations*" (IEA, 2002, p.40). Despite the need, research on this topic that is specific to demonstration and deployment remains sparse at best (Sagar and Gallagher, 2004). This research need can be further characterized as a lack of alignment in evaluation systems and lack of reliability testing for individual performance metrics.

The lack of common evaluation systems plays a role in the ability of any organization to iteratively evaluate the design and effectiveness of the organization in meeting stated goals (Saad, 2001). There is only a small amount of literature that addresses this topic in governmental AF deployment programs.. Howell and Chelius designed and executed a survey among state AF deployment programs to assess the effectiveness of differing policy mechanisms used by various state agencies (Howell and Chelius, 1997). In doing so, they found that many states lack common evaluation systems or often lack evaluation systems altogether (Howell and Chelius 1997). Many of the state agencies surveyed expressed little confidence in meeting the goals of the program. There is a dearth of literature or research that addresses this area despite the need and potential benefits of doing so.

In addition to the problem of commonality between programmatic AF deployment evaluation systems is the reliability of the metric itself. The performance evaluation literature deems the use of reliable, valid, and measurable performance metrics as critical to effectively assessing performance towards goals (Hurst, 1980; Smith, 1990; Kaplan and Norton, 1992; Neely, et al., 1995; Heimann, 1995; Kennerley and Neely, 2002; Behn, 2003). The metrics used to measure progress of AF deployment programs against goals are diverse and numerous. Little research addresses the reliability and validity of the performance measures. Howell and Chelius again represent one of the few studies in the literature that address this topic (Howell and Chelius, 1997). Their study on state AFV programs that attempt to deploy AFs and AFVs found that a lack of clear, measurable performance indicators were a major contributor ineffectiveness of performance measurement within these programs. This was found to be a primary factor in the lack of evaluative information needed to gauge progress towards goals within these programs.

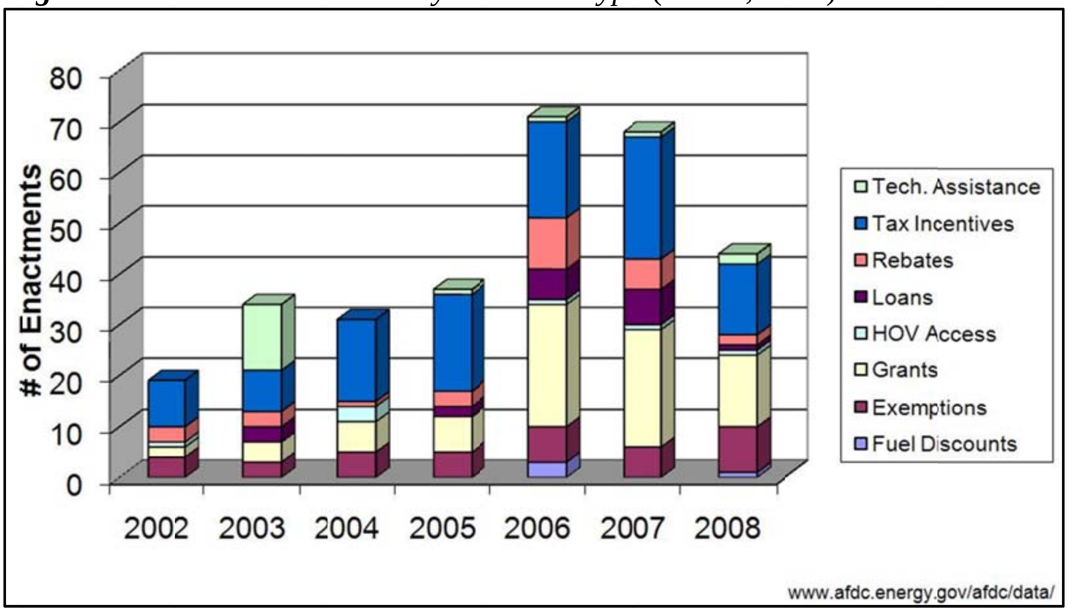
The lack of aligned goals, aligned evaluation systems, and reliable performance indicators presents more than just policy or program needs. There is a distinct methodological need as well. Methodologies that could be used to address these topics are as sparse as the research on the topics themselves. How is goal alignment between policies and programs measured? How can we evaluate the commonality of evaluation systems currently in place? How is the reliability and validity of individual performance indicators assessed? These questions may all fall under the broader concept of evaluation capacity building. The term evaluation capacity building (ECB) or building evaluation capacity (BEC) is a means for addressing both "inter" and "intra-organizational" goals and evaluation systems (Stockdill, et al., 2002, p.9). In general, ECB is defined as the "*intentional work to continuously create and sustain overall organizational processes that make quality evaluation and its uses routine*" (Stockdill, et al., 2002, p.14). In this research, ECB refers specifically to the work necessary to align, prioritize, and effectively evaluate progress towards goals associated with the deployment of AFs and AFVs.

Evaluation capacity building is a broad concept that has been applied to many fields. Clewell and Campbell utilized the concept of ECB towards a system of "cross-project evaluation" in which common goals and evaluations systems are “identified”, “refined”, and “operationalized” (Clewell and Campbell, 2008, p.4-6). The cross-project evaluation method was designed to better evaluate broad governmental and private programs (the National Science Foundation, the National Institute of Health, and the GE foundation) that were intended to increase the number of underrepresented groups in science, technology, engineering, and math (STEM) areas. Cross-project evaluation was needed because “*…several programs are funded that address the same or similar goals and the same or similar target populations*” (Clewell and Campbell, 2008, p.2). Therefore, “*the goal of cross-project evaluation is most often to provide data for a summative evaluation by assessing the program’s overall success in meeting its goals as measured by the success of individual projects in contributing to those goals*” (Clewell and Campbell, 2008, p.2). In doing so, it also becomes possible to then compare cumulative program effectiveness. The programmatic problems that ECB is used to address are nearly identical to those associated with AF deployment programs (previously described).

A primary barrier to researching ECB in the context of AF deployment programs is the existence of programs that state goals and evaluate progress against those goals in any kind of observable manner. As was stated by Howell and Chelius, most AF and AFV deployment programs “*…lack clear and quantifiable information on AFV goals and progress*” (Howell and Chelius, 1997, p.4). They further emphasize that “*this lack of information makes it difficult to evaluate the effectiveness of incentives in helping meet program goals*” (Howell and Chelius, 1997, p.4-5). However, Howell and Chelius go on to state that “*because states consider existing grants, rebates, and loans for AFV incremental costs or conversions to be the most important incentives, they are the primary focus for reporting on effectiveness of incentives*” (Howell and Chelius, 1997, p.5). Therefore from a methodological perspective these types of programs represent the few that are observable in terms of what the goals are and how evaluation is measured against these goals.

Grant programs that support the deployment of AFs are one of the most common and visible (by design) methods for supporting the AF demonstration and deployment. As testament to this fact, Figure 1 shows that grants represent approximately one-third of incentive-based enactments for deploying alternative fuel technologies as of 2008 (AFDC, 2010). Although on the decline, grants still represent over a fifth (21%) of all incentive types according to the AFDC in 2013 (AFDC, 2013).

***Figure 1**: Incentive Enactment by Incentive Type (AFDC, 2010)*



In addition to the frequency of using grants and their inherent visibility, AF grant programs (especially fuel-neutral ones) are a unique policy mechanism in that they are inherently one of the most observable and transparent. That is, they utilize a clearly scored system to choose which project or fuel type to support. The goals of these programs are not only clearly stated, but it is plainly clear which performance metrics are used to measure “success” towards these goals. This cannot be stated for any of the other types of incentives or enactments that are designed to aid deployment. In addition, they do not “pick winners” in the alternative fuel realm as they are fuel- and project-neutral (wide-scope). Therefore they cannot use performance measures that are fuel specific. This attribute is of particular interest because it is nearly a direct proxy for one of the most complex and vexing questions surrounding alternative fuel deployment: how do we choose which projects to fund among the many fuel types, vehicle platforms, and applications that are available? These programs are one of the only policy tools that is directly measurable in terms of understanding what we are using alternative fuels for and how aligned these policies are currently when considering an assessment of competing projects and fuel types.

For these reasons, grant programs are a prime candidate for building and demonstrating a methodology for building ECB in and across AF deployment programs. While the specific findings of goal alignment, evaluation alignment, and evaluation reliability may not be generalizable to all AF deployment programs, the methodology for *how* to implement the concept of ECB can be demonstrated. As a result, this research can serve as a foundation for how to implement ECB in other types of deployment programs as they advance and more evaluation opportunities become available.

POINT OF DEPARTURE

While literature routinely states the need for a more systematic and strategic method for identifying and aligning goals and evaluation procedures within and among AF demonstration and deployment programs, there is only sparse literature that develops and/or demonstrates a methodology for doing so. Additionally, there are none that describe these programs in the context of evaluation capacity. Therefore, the overarching point of departure of this research is to contribute to the body of AF deployment literature through building and demonstrating a strategic method for systematically evaluating goal alignment, evaluation alignment, and evaluation reliability under the broad heading of ECB. Rather than evaluating the current state of all AF deployment programs, this research develops and demonstrates how to build evaluation capacity in a single type of AF deployment programs: grant programs.

The first point of departure is the development of a methodology and framework for analyzing the alignment of goals and evaluative measures in alternative fuel deployment programs. This framework and methodology describes the current state of practice regarding programmatic goal and evaluation measure alignment. Having the ability to measure relative alignment and misalignment within specific goal domains provides the roadmap for where to study metric reliability in later sections. While this methodology and framework is only demonstrated on a specific portion of deployment programs (i.e. grant programs), this analysis could be applied to other types of deployment policies and programs. In addition, this framework and methodology can be applied to future and ongoing analysis of programmatic and policy alignment.

The second point of departure is the development of a methodology and framework for describing the reliability of individual evaluation criteria and performance metrics that are used to assess performance towards the stated goals of alternative fuel deployment programs, especially in areas of misalignment identified by the research pertaining to the first point of departure. The literature shows that there are myriad performance measures used to describe the effectiveness of alternative fuel deployment programs and that these criteria and measures *should* be compatible, consistent, and reliable (among several other listed attributes). The literature in this area has yet to define a methodology for describing the relative reliability of the evaluation criteria and performance indicators. This portion of the research creates and demonstrates a method for assessing the reliability of individual evaluation criteria and performance metrics.

The third and final point of departure is the development of a cumulative framework for building evaluation capacity within alternative fuel deployment programs. The relevant literature in this area routinely and consistently states the need for greater alignment, consistency, and better metrics to gauge progress; programmatic attributes that comprise evaluation capacity. However, there is no accepted means for building evaluation capacity in the arena of alternative fuel deployment. Therefore, this research fulfills a much greater gap in the literature by developing and demonstrating such a framework and does so under the broader heading of building evaluation capacity.

RESEARCH QUESTIONS

The overarching question that this research addresses is as follows:

> *How can the alignment of goals, and performance towards those goals, be cumulatively described in alternative fuel deployment programs?*

The first phase of this research program establishes a methodology for measuring the alignment of goals and evaluation systems used to measure performance against those goals among AF deployment grant programs. In doing so the current state of practice for this specific section of alternative fuel deployment programs is established. Thus the primary question that the first phase of research addresses is as follows:

What programmatic goals and evaluation systems exist and how well are they aligned among AF deployment grant programs?

The second phase of research addresses the reliability of the evaluation criteria and performance metrics currently used to describe the effectiveness of AF deployment grant programs. Thus the primary research question that this phase of research addresses is as follows:

How measurable and reliable are the criteria used to evaluate the energy security and technology transition performance of alternative fuel deployment programs?

The next sections detail the research tasks and methodologies associated with each of these research questions individually.

DISSERTATION FORMAT AND LAYOUT

This dissertation uses a non-traditional "three paper" dissertation format. This format is comprised of an opening chapter (this chapter) that summarizes the observed problems and theoretical points of departure addressed by the research contained in this dissertation. The next three chapters (Chapters 2, 3, and 4) each contain individual academic journal articles. Each chapter contains a separate abstract, introduction, methodology, set of results, conclusions and limitations as one might expect from individual peer-reviewed academic journals. Since multiple papers contained in the same research often share parts of methodology and background, there is overlap between each chapter, especially in the areas introduction and background areas that communicate both the research need and why grant programs are a unique domain for observing the selection process related to alternative fuel deployment projects.

At the time of this writing, chapter 2 was published in the December 2012 issue of the Journal of Energy Policy. The paper presented in chapter 3 will be submitted to the Journal of Transport Policy and the paper presented in chapter 4 will be submitted to the Journal of Energy Policy. Finally, Chapter 5 revisits the research questions and points of departure and presents a summary of the theoretical and practical contributions towards those questions and points.

CHAPTER REFERENCES

Alternative Fuels Data Center (AFDC), 2010. Data, Analysis, and Trends. United States Department of Energy, Alternative Fuels Data Center (online), retrieved February 17, 2010 from http://www.afdc.energy.gov/afdc/data/vehicles.html.

Alternative Fuels Data Center (AFDC), 2013. All Laws and Incentives Sorted by Type. United States Department of Energy, Alternative Fuels Data Center (online), retrieved June 14, 2013 from http://www.afdc.energy.gov/laws/matrix/incentive.

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

Bureau of Transportation Statistics (BTS), 2011. National Transportation Statistics 2011, United States Department of Transportation, BTS (online), retrieved July 7, 2011 from http://www.bts.gov/publications/national_transportation_statistics/pdf/entire.pdf.

Burns, T., Stalker, G.M., 1961. The Management of Innovation, Tavistock, London, UK.

Byrne, M.R., Polonsky, M.J., 2001. Impediments to Consumer Adoption of Sustainable Transportation – Alternative Fuel Vehicles. International Journal of Operations and Production Management, 21(12), 1521-1538.

Chun, Y.H., Rainey, H.G., 2005. Goal Ambiguity and Organizational Performance in U.S. Federal Agencies. Journal of Public Administration Research Theory, 15(1), 1-30.

Clewell, B.C., Campbell, P.B., 2008. Building Evaluation Capacity: Designing a Cross-Project Evaluation (Guide 1). The Urban Institute, January 2008, available online at http://www.urban.org/UploadedPDF/411651_guide1.pdf.

Deutch, J.M., 2011. An Energy Technology Corporation Will Improve the Federal Government's Efforts to Accelerate Energy Innovation, Brookings Institution, Discussion Paper 2011-05, May 2011.

Gallagher, K.S., Collantes, G., Holdren, J.P., Lee, H., Frosch, R., 2007. Policy Options for Reducing Oil Consumption and Greenhouse-Gas Emissions from the U.S. Transportation Sector, Energy Technology Innovation Policy, Discussion Paper, July 2007.

Heimann, L., 1995. Different Paths to Success: A Theory of Organizational Decision Making and Administrative Reliability. Journal of Public Administration Research and Theory, 5(1), 45-72.

Higgins, K.L., Maciariello, J.A., 2004. Leading Complex Collaboration in Network Organizations: A Multi-Disciplinary Approach, in: Beyerlein, M.M., Johnson, D.A., Beyerlein, S.T., (Eds.), Complex Collaboration: Building the Capabilities for Working Across Boundaries, Advances in Interdisciplinary Studies of Work Teams, Vol. 10, Elsevier, Ltd., Oxford, UK.

Howell, D. and Chelius, M., 1997. Alternative Fuel Vehicle Programs – Applicability of Government Incentives. Transportation Research Record, Volume 1587, 1-9.

Hurst, E.G., 1980. Attributes of Performance Measures. Public Productivity Review, 4(1), 43-49.

International Energy Agency (IEA), 2002. Deployment Strategies for Hybrid, Electric, and Alternative Fuel Vehicles, Annex VIII Final Report.

International Energy Agency (IEA), 2011. Technology Roadmap – Electric and Plug-In Hybrid Electric Vehicles, Updated June 2011.

Jung, C.S., 2011. Developing and Validating New Concepts and Measures of Program Goal Ambiguity in the U.S. Federal Government. Administration and Society, (online before print), 1-27.

Kaplan, R.S., Norton, D.P., 1992. The Balanced Scorecard – Measures that Drive Performance. Harvard Business Review, Jan/Feb 1992, 1-10.

Kennerley, M., Neely, A., 2002. A Framework of the Factors Affecting the Evolution of Performance Measurement Systems. International Journal of Operations and Production Management, 22(11), 1222-1245.

Lawrence, P.R., Lorsch, J.W., 1967. Organization and Environment: Managing Differentiation and Integration. Harvard University Press, Boston, MA.

Lee, H., 2009. Oil Security and the Transportation Sector, in: Gallagher, K.S. (Ed.), Acting in Time on Energy Policy. Brookings Institution Press, Washington, D.C.

Locke, E.A., 1968. Toward A Theory of Task Motivation and Incentives. Organizational Behavior and Human Performance, 3, 157-189.

Melaina, M., McQueen, S., Brinch, J., 2008. Refueling Infrastructure for Alternative Fuel Vehicles: Lessons Learned for Hydrogen – Workshop Proceedings, National Renewable Energy Laboratory (NREL), Proceedings NREL/BK-560-43669, July 2008.

Melendez, M., 2006. Transitioning to a Hydrogen Future: Learning from the Alternative Fuels Experience, National Renewable Energy Laboratory (NREL), Technical Report NREL/TP-540-39423, February 2006.

Miles, R.E., Snow, C.C., Meyer, A.D., Coleman, H.J., 1978. Organizational Strategy, Structure, and Process. The Academy of Management Review, 3(3), pp.546-562.

Miller, D., Friesen, P., 1984. A Longitudinal Study of the Corporate Life Cycle. Management Science, 30, 1161-1183.

Mintzberg, H., 1979. The Structuring of Organizations, Prentice-Hall, Englewood Cliffs, NJ.

Moynihan, D.P., 2008. The Dynamics of Performance Management: Constructing Information and Reform, Georgetown University Press, Washington, DC.

National Renewable Energy Laboratory (NREL), 2011. Commercialization and Deployment at NREL – Advancing Renewable Energy and Energy Efficiency at Speed and Scale, NREL/MP-6A42-51947, May 2011.

Neely, A., Gregory, M., Platts, K., 1995. Performance Measurement System Design: A Literature Review and Research Agenda. International Journal of Operations & Production Management, 15(4), 80 – 116.

Ouchi, W.G., 1980. Markets, Bureaucracies, and Clans. Administrative Science Quarterly, 25(1), pp.129-141.
Rainey, H.G., 2009. Understanding and Managing Public Organizations, 4th ed., Jossey-Bass, San Francisco, CA.
Saad. G.H., 2001. Strategic Performance Evaluation: Descriptive and Prescriptive Analysis. Industrial Management and Data Systems, 101(8), 390-399.
Sagar, A., Gallagher, K.S., 2004. Energy Technology Demonstration and Deployment, in: National Commission on Energy Policy (NCEP) Technical Appendix for Chapter 6: Developing Better Energy Technologies for the Future, in: Ending the Energy Stalemate – A Bipartisan Strategy to Meet America's Energy Challenges, December 2004.
Smith, P., 1990. The Use of Performance Indicators in the Public Sector. Journal of the Royal Statistical Society, 153(1), 53-72.
Sperling, D., Gordon, D., 2009. Two Billion Cars: Driving Towards Sustainability, Oxford University Press, New York, New York.
Sperling, D., Yeh, S., 2009. Low Carbon Fuel Standards. Issues in Science and Technology (online), Winter 2009, retrieved June 16, 2011 from http://www.issues.org/25.2/sperling.html.
Sperling, D., 1988. New Transportation Fuels – A Strategic Approach to Technological Change, University of California Press, Berkeley, California.
Stockdill, S.H., Baizerman, M., Compton, D.W., 2002. Toward a Definition of the ECB process: A Conversation with the ECB literature, in: Compton , D.W., Baizerman, M., Stockdill, S., (eds.), The Art, Craft, and Science of Evaluation Capacity Building, Vol. 93, New Directions for Evaluation, Jossey-Bass, San Francisco, CA.
United States Congress (USC), 1993. Government Results and Performance Act, Public Law 103-62, 103rd Congress.
United States Department of Energy (USDOE), 2011. Federal Energy Management Program: Laws and Regulations (online), retrieved November 7, 2011 from http://www1.eere.energy.gov/femp/regulations/regulations.html.
United States Energy Information Administration (USEIA), 2011. Alternatives to Traditional Transportation Fuels 2009, Released April 2011, USEIA (online), retrieved July 7, 2011 from http://www.eia.gov/renewable/alternative_transport_vehicles/pdf/afv-atf2009.pdf.
United States Energy Information Administration (USEIA), 2013. How many alternative fuel and hybrid vehicles are there in the U.S.?, USEIA (online), retrieved June 14, 2013 from http://www.eia.gov/tools/faqs/faq.cfm?id=93&t=4.
United States Government Accountability Office (USGAO), 2011. Opportunities to Reduce Potential Duplication in Government Programs, Save Tax Dollars, and Enhance Revenue. Report GAO 11-318SP, March 2011.
Witt, L.A., 1998. Enhancing Organizational Goal Congruence: A Solution to Organizational Politics. Journal of Applied Psychology, 83(4), 666-674.
Woodward, J., 1965. Industrial Organization: Theory and Practice, Oxford University Press, London, UK.

CHAPTER 2: MAPPING GOAL ALIGNMENT OF DEPLOYMENT PROGRAMS FOR ALTERNATIVE FUEL TECHNOLOGIES: AN ANALYSIS OF WIDE-SCOPE GRANT PROGRAMS IN THE UNITED STATES

ABSTRACT

Governments have attempted to advance alternative fuels (AFs) in the on-road transportation sector with the goal of addressing multiple environmental, energy security, economic growth, and technology transition objectives. However there is little agreement, at all governmental levels, on how to prioritize goals and how to measure progress towards goals. Literature suggests that a consistent, aligned, and prioritized approach will increase the effectiveness of deployment efforts. While literature states that goal alignment and prioritization *should* occur, there are few studies suggesting *how* to measure the alignment of deployment programs. This paper presents a methodology for measuring goal alignment by applying the theories of goal ambiguity. It then demonstrates this methodology within the context of fuel- and project-neutral (wide-scope) grant programs directed toward AF deployment. This paper analyzes forty-seven (47) wide-scope federal, state, and regional grant programs in the United States, active between 2006 and 2011. On the whole, governments most use deployment grant programs to address environmental concerns and are highly aligned in doing so between agency levels. In contrast, there is much less consensus (and therefore goal alignment) on whether or how governments should address other priorities such as energy security, economic growth, and technology transition.

INTRODUCTION

Deployment of AF technologies in the on-road transportation sector often attempts to address multiple objectives. These can include accelerating technology transition, lessening environmental impact, enhancing energy security, and spurring economic growth through job and industry creation. Yet as various levels of government in the United States devise programs to deploy AF technologies to address any one or combination of these goals, there is little consensus on the goals to be addressed, their prioritization, or how to effectively measure success against the objective(s). Despite considerable and sustained deployment efforts over the past 40 years, alternative fuel vehicles (AFVs) represent less than one percent of the light and heavy-duty vehicles on the road today (BTS, 2011; USEIA, 2011). This is at least in part a consequence of AF program designs that overlook the benefits of strategic goal alignment between programs. Consistent, aligned and prioritized policy goals and technology deployment designs offer opportunities to reduce or eliminate conflicting and counterproductive programmatic elements and capture system-wide synergies available through alignment of common objectives.

Literature routinely states that consistent, aligned and prioritized policies and messages from governments at state, regional, federal, and even international levels can improve current and future efforts to deploy AF technologies (USGAO, 2011; Deutch, 2011; IEA, 2011; NREL, 2011; Sperling and Gordon, 2009; Sperling and Yeh, 2009; Lee, 2009; Melaina et al., 2008; Melendez, 2006; Sagar and Gallagher, 2004; Byrne and Polonsky, 2001; IEA, 2002; Howell and Chelius,1997; Sperling, 1988). Yet there is a dearth of research that specifically addresses technology deployment efforts (Sagar and Gallagher, 2004) and even fewer that address strategic goal alignment between programs, or the consistency of criteria used to measure "success" against these goals. Consequently, there are few descriptions of how aligned deployment programs currently are, or even how to measure alignment.

The need for increased alignment and coherence in goals is neither new nor unique to the AF sector. Low levels of coherence are often attributed to poor performance, whether by a lack of consistency in policy construction or policy implementation (Mazmanian and Sabatier, 1983).

Alternative fuel policies governing the transportation sector are no exception. Arguably, they are even more prone to misalignment because policymakers are attempting to address a number of objectives that span multiple domains (e.g. environment, national security, and economic growth) (Deutch, 2011; Sperling and Gordon, 2009; Sperling, 1988). Deutch emphasizes that, "…a single objective—for example, reducing emissions—is simplest, but multiple objectives are the rule…" (Deutch, 2011). In addition, there is an extremely diverse group of stakeholders involved in AF deployment, with multiple needs and objectives and with differing means for expressing them (Byrne and Polonsky, 2001; Sperling and Gordon, 2009). Consequently, AF policies end up incorporating a patchwork of intentions and goals expressed in a variety of ways that are occasionally divergent, if not entirely conflicting.

Questions concerning prioritization, alignment, and coherence of goals, as well as measuring performance against these goals, may all fall under the broader heading of "goal ambiguity." Goal ambiguity is generally defined as the latitude that people or organizations are allowed in defining goals and measuring performance against those goals (adapted from multiple definitions given by Chun and Rainey, 2005). Goal ambiguity is a subset of goal setting theory as defined by Locke (1968), where higher and more clearly defined goals result in higher performance in terms of human motivation. Goal ambiguity theory helps explain why and how higher performance is achieved as measured by goal-setting, when higher or lower degrees of variability in definitions, directives, performance indicators, and evaluation criteria are present (Jung, 2011; Chun and Rainey, 2005; Latham and Locke, 1991; Matland, 1995; Ripley and Franklin, 1982). Like goal setting theory, previous studies conclude that decreased goal ambiguity may lead to more efficient policy outcomes (Matland, 1995; Ripley and Franklin, 1982). Consequently, the concept of goal ambiguity has been applied to many areas of policy analysis including healthcare organizations (Calciolari et al., 2011), human service agencies (Pandey and Wright, 2006), and federal agency programs (Jung, 2011; Chun and Rainey, 2005), to name a few. In this paper we apply the theory and practical concepts of goal ambiguity to analyze efforts designed to deploy AFs in the on-road transportation sector.

Chun and Rainey distinguish four unique categories of goal ambiguity:

- "Mission comprehension ambiguity;
- Directive goal ambiguity;
- Evaluative goal ambiguity; and
- Priority goal ambiguity" (Chun and Rainey, 2005, p.529).

Mission comprehension and directive goal ambiguities address the *stated* purposes and actions of organizations. For technology deployment programs, funding is often distributed in a manner that is not amenable to measurement against stated objectives (IEA, 2002). Consequently, an analysis of stated programmatic goals often does not necessarily reflect reality and we do not address these areas of goal ambiguity. Conversely, priority and evaluative categories of goal ambiguity relate directly to the ordering of objectives and the measures organizations use to gauge progress against them. Hence, these are the areas of need most expressed in literature and reflect the areas of goal ambiguity we seek to address. In this paper, priority goal ambiguity is defined as the degree of variability in specified goals between agencies and organizations that seek to deploy AFs. Evaluative goal ambiguity is defined as the degree of variability in how organizations or agencies measure performance against goals.

An important discussion topic on goal ambiguity is whether higher or lower levels of goal ambiguity are a positive or negative attribute. As was previously presented, much of the literature on AF deployment broadly suggests that consistent messages and greater alignment (less ambiguity) is a positive policy and program attribute. However, the literature does not typically distinguish between priority and evaluative ambiguity as we do in this paper. There are several areas of research that suggest greater degrees of policy efficiency and effectiveness may be realized if different levels of government are able to address different areas of policy. This is because different levels of government may be more advantageously structured to address particular problems. For example, research under the headings of fiscal federalism (Sharma, 2011), environmental federalism (Banzhaf and Chupp, 2010; Vogel, et al., 2010), polycentrism (Sovacool, 2011), and top-down versus bottom-up approaches (Lutsey and Sperling, 2008), to name a few of the many headings, all speak to a common question: are centralized or decentralized approaches more or less effective and efficient in accomplishing goals? This concept is particularly relevant in the area of priority goal ambiguity. That is, high or low priority goal ambiguity may or may not be a positive attribute. For this research, comparison of the effectiveness or efficiency of a centralized versus a decentralized goal approach across governmental agency levels is beyond the scope of this paper. The primary contribution of this paper is to present a systematic methodology for determining and tracking the level of priority goal ambiguity in AF deployment programs. Therefore differing levels of priority goal ambiguity are presented as neither a positive nor a negative attribute of AF deployment programs. Instead, varying levels of priority goal ambiguity are presented in this paper as a current state of practice for determining what goals are currently addressed by AF deployment programs and by which agency level.

The same interpretation of higher or lower levels of goal ambiguity does not apply to the area of evaluative goal ambiguity. Higher levels of evaluative goal ambiguity indicate that, when common goals exist, common metrics are not used to track progress towards those goals. Commonality of evaluation systems and the need for greater, common knowledge (less evaluative goal ambiguity) is a routinely stated need in AF literature (NREL, 2011; Sperling and Gordon, 2009; IEA, 2002). Therefore we present lower levels of evaluative goal ambiguity as a positive attribute in this analysis. In this paper we investigate evaluative goal ambiguity across goal domains to provide a high level view of the disparity in how progress is measured within each goal domain.

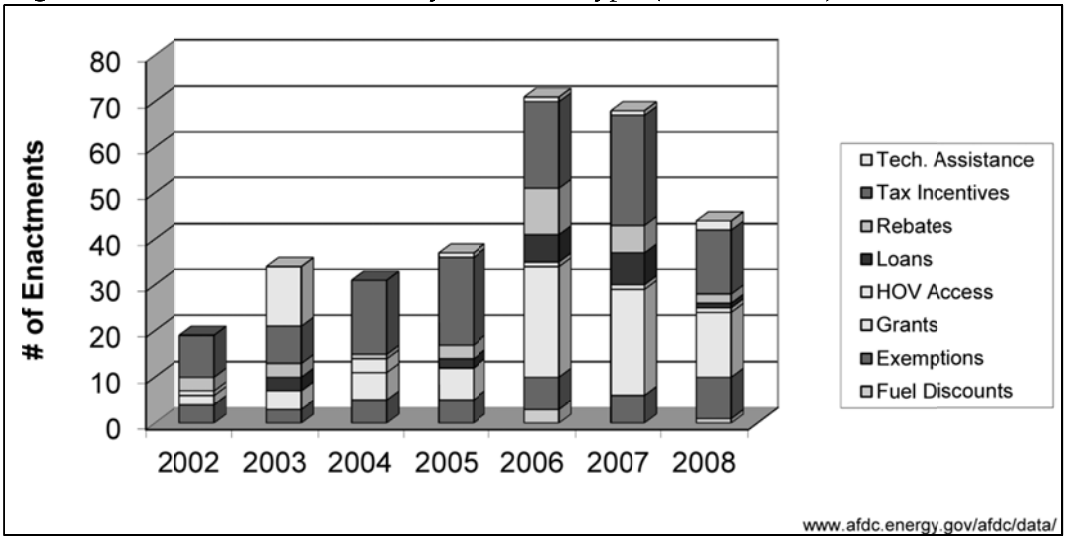
WIDE-SCOPE GRANT PROGRAMS – AN OPPORTUNITY FOR DEPLOYMENT GOAL RESEARCH

Policy coherence or alignment is often described as a need, but is often elusive in terms of discernibility, measurement, and interpretation (May et al., 2006). Specifically, the lack of discernibility (the ability to observe specific intent or goals of a policy mechanism) is common within many of the policy mechanisms used to aid the deployment of AFs (IEA, 2002). As a result, identifying the goals or measurement systems used to assess progress against these goals can prove to be quite difficult.

In contrast, AF grant programs that utilize a competitive selection process for distributing funds are a unique policy mechanism in that they are *observable* and *transparent* by design. That is, they utilize a clearly scored system to choose which project or fuel type to support. The goals of these programs are clearly articulated through evaluation criteria and performance metrics used to measure “success” against these goals. In addition, grant programs commonly attempt to avoid “picking winners” in the AF arena by implementing fuel-neutral criteria and metrics. This proscribes the use of performance measures that are fuel specific. Consequently, AF technology deployment grant programs are one of the few policy mechanisms that can be characterized as directly observable and measurable with regard to their goal definition, prioritization, and alignment.

In addition to the inherent discernibility of grant programs, frequency of use is similarly well described. Grant programs are one of the most common methods for supporting the deployment of AFs. These programs are administered by many federal, state, and regional agencies including the Department of Energy (DOE), the Federal Highway Administration (FHWA), numerous state energy offices (SEOs), and regional agencies such as councils of government and metropolitan planning organizations (MPOs). Figure 1 shows that grants represent approximately one-third of incentive-based enactments for deploying AF technologies as of 2008 (AFDC, 2010). Consequently, improvement to one of the more commonly used AF policy types will likely enhance the efficiency of incentive enactments more broadly. Similarly, the research methodology applied to the grant-based policy type may offer insight for research into other policy mechanisms as well.

***Figure 1**: Incentive Enactment by Incentive Type* (*AFDC, 2010*)



Note: Incentive types in the bar chart are shown in the order given in the legend.

RESEARCH APPROACH

For this research, the grant program (whether at the federal, state, or regional level) is the unit of analysis. A request for proposal (RFP) is the most common instrument used by advertising agencies for soliciting grant proposals and is the unit of observation. Grant programs often solicit proposals on an annual basis. To avoid over-representing any one program, the research analyzes only one RFP per grant program. This ensures that any program extending over consecutive years is not repeatedly analyzed over that time period. A content analysis methodology was chosen for analyzing the RFP documents. A content analysis is generally defined as, "…any technique for making inferences by objectively and systematically identifying specified characteristics of messages" (Holsti, 1969, p.14). In this application, the content analysis methodology serves as a quantitative means for measuring the usage frequency, performance indicators, and consequently, the goals of the RFP documents analyzed. The process of defining coding categories in which frequency is measured, is "…a combination of induction and deduction…" in which *a priori* categories are defined, but augmented with iteratively defined categories as the content analysis process progresses and data points are evaluated (Neuendorf, 2002, p.11-12). The content analysis process of evaluation consists of four basic steps (Neuendorf, 2002):

1. Creation of content categories and subcategories;
2. Coding (of the solicitation documents) within the defined categories and subcategories;
3. Analysis of results defined by frequency of categories and subcategories usage; and
4. Analysis of results by explanatory variables.

To be included in this study, a grant program must:

- Allow multiple AFs to compete for selection under the same grant solicitation;
- Include at least one combustible AF (e.g. ethanol, compressed natural gas (CNG), liquefied natural gas (LNG), liquefied petroleum gas (LPG), biodiesel) as an eligible fuel;
- Include at least one zero tailpipe emission (ZTE) fuel (e.g. hydrogen or electricity) as an eligible fuel;
- Target on-road AF transportation projects; and
- Support alternative fuel vehicle (AFV) purchases, construction of alternative refueling infrastructure, or both.

For the purposes of this study a broad definition of AFs was used, based upon the definition provided by the 1992 Environmental Policy Act (EPAct) which includes methanol, ethanol, biodiesel, CNG, LNG, LPG, hydrogen, electricity, and advanced biofuels (USC, 1992).

The population of grant programs that fit these membership criteria is not easily defined since there is no single database where all such programs are listed. The Alternative Fuels Data Center

(AFDC) and the Database of State Incentives for Renewable Energy (DSIRE) provide some guidance for finding such programs (AFDC, 2011; USDOE, 2011). According to these resources there are 7 federal and 24 state programs for which RFPs may be available that fit the membership rules previously defined. Within these databases, 6 RFPs were found for federal programs representing 86 percent of the possible programs listed. At the state level, RFPs were found for 17 of the 24 listed programs, representing 71 percent of the listed eligible programs. However, internet resources and agency solicitation websites yielded 21 programs not specifically listed in either database. Consequently, the population of programs or available RFPs from these programs cannot be clearly defined. In total, 47 grant programs were included in the analysis, representing 25 states over a period from 2006 to 2011.

Representative data points (grant programs analyzed) are classified by the advertising agency. The sample includes federal, state, and regional advertising agencies. Tables 1 through 3 list the evaluated grant programs by proposal year, program dollar value, and targeted project scope (e.g. AFV purchase, refueling infrastructure construction, or both). Notably, the 2009-2010 period comprises the bulk of grant programs analyzed. However, several programs do not solicit proposals every year and did not solicit proposals during the 2009 to 2010 period. In efforts to construct a requisite data set, RFPs from years adjacent to the 2009 to 2010 period were included as well.

Table 1: *Programs Analyzed by Advertising Agency and Year*

Advertising Agency / Grant Year	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**Σ**
Federal	0	1	0	2	3	0	**6**
State	1	1	2	7	12	3	**26**
Regional	0	0	0	4	11	0	**15**
Total	**1**	**2**	**2**	**13**	**26**	**3**	**47**

Table 2: *Programs Analyzed by Advertising Agency and Grant Value*

Advertising Agency / Grant Value	**Not Specified**	**< $1M**	**$1M to 10M**	**>$10M**	**Σ**
Federal	0	0	1	5	**6**
State	7	3	11	5	**26**
Regional	4	4	4	3	**15**
Total	**11**	**7**	**16**	**13**	**47**

Table 3: *Programs Analyzed by Advertising Agency and Targeted Scope*

Advertising Agency / Targeted Scope	**Vehicles Only**	**Refueling Infr. Only**	**Both**	**Σ**
Federal	3	0	3	**6**
State	9	1	16	**26**
Regional	5	2	8	**15**
Total	**17**	**3**	**27**	**47**

RESULTS

The evaluation criteria identified in each RFP were categorized into four general goal domains: environment, energy security, economic growth, and technology transition. Evaluation criteria were categorized into goal domains according to what the criteria directly measure. For example, criterion measurement units of tons of emissions reduced (e.g. NOx or SOx for example) were assigned to the "environment" goal domain category because these criteria directly measure progress towards environmental goals. Notably, there is some subjectivity in these assignments. This criticism is common for the methodology employed in this research. To address this, the categorization system was reviewed by agency officials at different levels for adequacy and appositeness. In addition, we present the relative frequency results for each criterion in the following table. The reader is invited to reassign the criteria to different categories if they so choose.

The results are shown in Table 4 and are presented by advertising agency. Notably, more evaluation criteria were discovered than are displayed in Table 4. However, other evaluation criteria were not goal-specific, and typically came in the form of project viability criteria (e.g. viability of construction schedule, amount of team experience, etc.). We chose to consider these criteria as necessary for achieving goals representative of all goal domains. Consequently, they were not included in this analysis. The following sections address the varying levels of priority goal ambiguity and evaluative goal ambiguity. Results for both types of ambiguity are presented from both a combined perspective (e.g. federal, state, and regional programs together) and individually.

PRIORITY GOAL AMBIGUITY

In this paper, priority goal ambiguity is defined as the degree of variability in specified goals between agencies and organizations that seek to deploy AFs. These differences are presented in the context of four goal domains: environment, economic growth, energy security, and technology transition. Representation in a goal domain is measured by the number of times that goal-specific, evaluative measures appear (at least once) in the RFP document evaluated. Individual agency level results (federal, state, *or* regional) are calculated by dividing the maximum number of times any single criterion within a goal domain was cited, by the total number of proposals analyzed at that agency level. This analysis provides a high level view of which goal domains are most addressed (goal prioritization) by agencies at different governmental levels. Combined results (federal, state, *and* regional) are also presented. To construct the combined results, the maximum number of times any single criterion was cited within a goal domain at each agency level is summed across all agency levels and divided by the total number of projects analyzed at all agency levels. This analysis provides a high level view of which goal domains are addressed (goal prioritization) by AF deployment grant programs in general. These results are presented in Figures 2 through 5.

When looking at the combined results of all programs analyzed, it is clear that environmental goals are most commonly addressed and were found in 85 percent of the grant programs analyzed (see Figure 2). Conversely, energy security, economic growth, and technology transition goals were found in only 36 to 43 percent of the programs analyzed. On the whole, it can be stated that environmental goals are the most commonly cited objective of AF deployment grant programs. While the other goal domains are less represented, they are represented at approximately the same level in comparison to each other. However, when looking at criterion representation for federal, state, or regional grant programs individually (see Figures 3-5), the results clearly indicate that there are major variances in the representation of goals in which AF technology deployment is advanced as a solution.

At all agency levels (see Figures 3-5), reducing environmental impact proved the most prevalent goal, appearing the most frequently in solicitation documents, and with approximately the same frequency for each agency category. This indicates that deployment grant programs are most commonly used to address environmental goals at all individual agency/organization category levels. In addition, this indicates that there is a very low degree of priority goal ambiguity when considering this goal domain. This finding is congruent with other studies that have also identified environmental goals as the primary purpose for developing and distributing alternative energy sources (Shen, et al., 2010). Conversely, there is a great deal of disparity in how often AF deployment grants are used to address economic, energy security, and technology transition goals.

Table 4: *Number of Grant Programs Citing a Specific Criterion (at Least Once)*

Goal Domain	Objective	Evaluation Criteria	(Federal) [6 Programs]	(State) [26 Programs]	(Regional) [15 Programs]
Environment	Air Quality Maximized	Air Pollution Emission Rates (CO, PM, NO_x, and SO_x)	5 (83%)	23 (88%)	12 (80%)
	Env. Impact Minimized	GHG Emission Rates (CO_2, Methane, and N_2O)	2 (33%)	12 (46%)	5 (33%)
		Water Pollution Rates (*from fuel use/ feedstock production*)	0	3 (12%)	0
		Land Use Impact (*from feedstock production*)	0	2 (8%)	0
		Water Use Impact (*from feedstock production*)	0	3 (12%)	0
Econ. Growth	Econ. Stimulus Maximized	Immediacy of Project Implementation	4 (67%)	2 (8%)	5 (33%)
		# of Jobs Created/Saved (Quantified)	1 (17%)	8 (31%)	3 (20%)
		Discussion of Jobs Created/Saved (Qualified)	1 (17%)	6 (23%)	1 (7%)
	GDP Maximized / Mfr. Dev.	New Mfrs./Companies Created	1 (17%)	9 (35%)	3 (20%)
		Dev./Maint. of Existing Mfrs./Companies	1 (17%)	11 (42%)	3 (20%)
Energy Security	Dependence on Trad. Energy Minimized	# of Gallons of Petroleum Displaced	1 (17%)	13 (50%)	3 (20%)
Technology Transition	Dist. Of AFVs Maximized	# of AFVs Purchased/Subsidized	3 (50%)	9 (35%)	4 (27%)
		Ratio of AFVs to Traditionally Fueled Vehicles	0 (0%)	0 (0%)	1 (7%)
		Density of AFVs in a Geographic Area	2 (33%)	0 (0%)	0 (0%)
		# of AFVs by Vehicle Use	1 (17%)	4 (15%)	0 (0%)
		# of AFVs by Weight Class	0 (0%)	2 (8%)	2 (13%)
		# of AFVs by Targeted Corridor	0 (0%)	1 (4%)	0 (0%)
	Usage of AFVs Maximized	VMT Traveled with an AFV	0 (0%)	7 (27%)	3 (20%)
		# of Hours the AFV is Used	0 (0%)	2 (8%)	0 (0%)
		# of Years the AFV will be Used	1 (17%)	5 (19%)	1 (7%)
		Ridership/Tons of Freight Conveyed with an AFV	0 (0%)	3 (12%)	3 (20%)
	Dist. Of AFs Maximized	# of GGE Distributed	1 (17%)	4 (15%)	2 (13%)
		# of AF Refueling Stations Constructed	0 (0%)	3 (12%)	0 (0%)
		# of Customers Served	1 (17%)	0 (0%)	0 (0%)
		Discussion of Market Population	0 (0%)	5 (19%)	1 (7%)
		Public Access to AF Refueling Stations	1 (17%)	5 (19%)	3 (20%)
		Geographical Need for AF Refueling Station	3 (50%)	2 (8%)	1 (7%)
	AF Market Transformation/ Acceptance Maximized	Signage for Location/Price of AF Refueling Station	1 (17%)	3 (12%)	4 (27%)
		# of Outreach/Educational Events	3 (50%)	7 (27%)	4 (27%)
		Discussion of Contrib. to Future Market Strength	3 (50%)	11 (42%)	4 (27%)

***Figure 2**: Representation of Goals Addressed by Federal, State, and Regional AF Deployment Grant Programs*

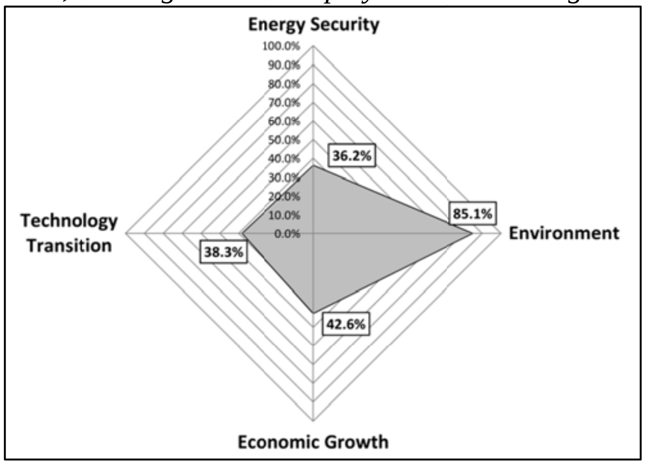


***Figure 3**: Representation of Goals Addressed by Federal AF Deployment Grant Programs*

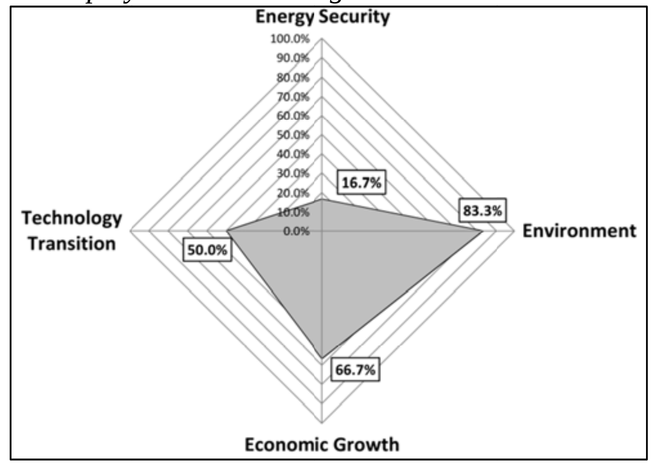


***Figure 4**: Representation of Goals Addressed by State AF Deployment Grant Programs*

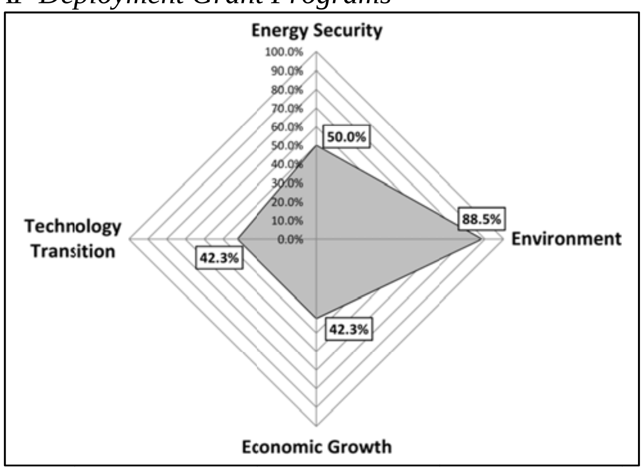
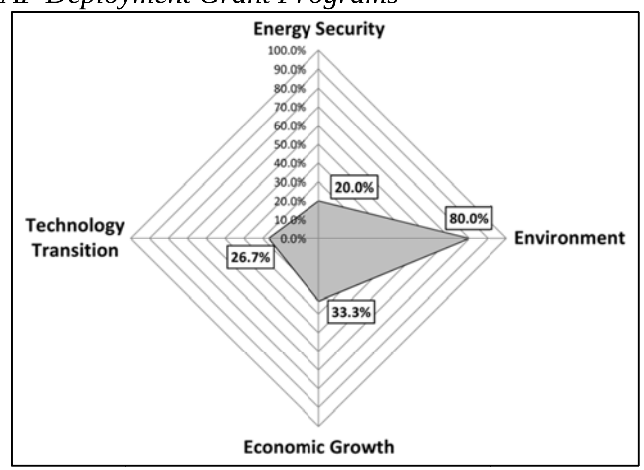


***Figure 5**: Representation of Goals Addressed by Regional AF Deployment Grant Programs*



Economic growth goals were assessed by the inclusion of traditional evaluation criteria (jobs created, jobs saved, new businesses created, etc.) found in the RFP solicitation documents. Notably, these evaluation criteria are not unique to the realm of AFs, since the same criteria could be used to assess the performance of any project or program designed to provide economic stimulus. Federal agencies appear to favor the deployment of AFs for addressing economic goals. Of the criteria used to evaluate the economic contribution of these projects, the "immediacy of project implementation" criterion was most commonly used. This is likely a direct effect of the sizeable 2009 American Recovery and Reinvestment Act (ARRA) legislation that funded many alternative energy projects (USC, 2009). In contrast to the federal level of funding, state and regional agencies appear to address traditional economic goals with much less frequency (42 and 33 percent of programs analyzed, respectively). In addition, state and regional agencies appear to do so with approximately the same frequency. These findings indicate that there is some disconnect (at least during the period of funding this research addresses) between the goals of federal agencies and the goals of state and regional agencies when considering the goal domain of economic growth. Irrespective of any effect on economic growth, the programs analyzed in this research display a significant degree of misalignment within this goal domain.

Programs commonly addressed energy security as a goal, but with a high degree of priority goal ambiguity between the federal, state, and regional agency/organization categories (17, 50, and 20 percent of projects analyzed respectively). Interestingly, state programs appear to strongly favor the goal of energy security over federal and regional programs. We speculate that uncertainty over a common definition of energy security impedes the inclusion of this goal into federal policy. At the federal level, definition and inclusion of measures towards goals can be an arduous and convoluted process due to the contentious atmosphere of lawmaking surrounding renewable energy. This phenomenon has generally resulted in federal policies that address energy security less robustly, even though the concept is politically popular (Bang, 2010). Conversely, state level governments are often more capable of efficiency in adopting measures due to increased levels of political cohesiveness. This may have resulted in increased representation at the state level.

A caveat to these findings is the definition of energy security used in this research. Previous and current research on energy security definitions and indicators suggests continuing and broad disparity in how energy security is defined and measured (see Kruyt, et al., 2009; Chester, 2010; Loschel, et al., 2010; Jansen and Seebregts, 2010; Greene, 2010; Winzer, 2012). In the categorization system used in this research only one evaluation criterion represents the energy security goal domain. This is because a search of the solicitation documents located only one measurable proxy that directly addresses energy security: a reduction in the consumption of petroleum-based (gasoline and diesel) fuels. While the search yielded other evaluation criteria that indirectly measure the amount of petroleum-based fuels displaced, (e.g. number of AFVs distributed, VMT traveled with an AFV, etc.), these evaluation criteria do not directly measure the amount of petroleum displaced. Further, these more directly and quantitatively measure success against other goal domains (e.g. technology transition from conventional vehicles and fuels to AFVs or AFs). However, the use of alternative definitions of energy security that do not distinguish between environmental, technology transition, and energy security goal domains as we have, could yield alternative results.

When considering the technology transition goal domain, the research uncovered some degree of priority goal ambiguity between federal, state, and regional agency categories (50, 42, and 27 percent of projects analyzed respectively). While federal and state agencies appear to prioritize deployment as an important goal (evaluated approximately half of the time), the findings suggest that regional agencies do so to a lesser degree (addressed approximately one-third of the time). We speculate that the construction of grant programs designed to decrease regional environmental impacts may contribute to this outcome. Within such programs environmental goals are often the primary purpose of the grant, with AF deployment used as a means for addressing them. However, the benefits of a systematic technology transition effort towards the use of AFs are not an expressed goal when perhaps they should be.

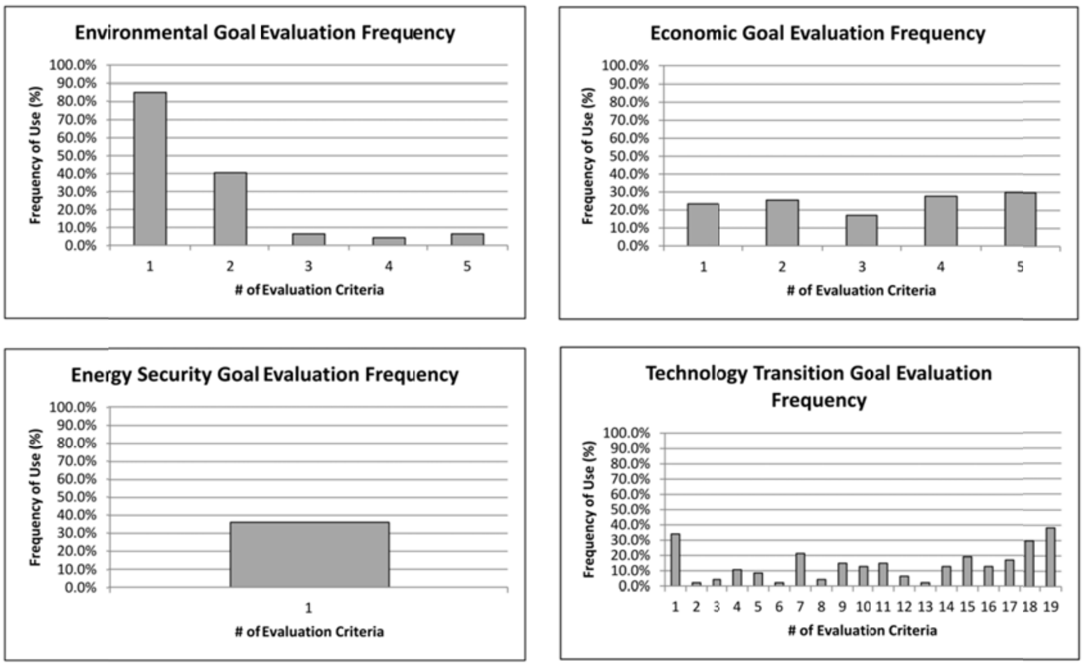
EVALUATIVE GOAL AMBIGUITY

While the frequency with which goals are addressed provides one measure of goal alignment between programs, the method and means of measuring performance towards those goals provides another. In this paper we define this property as evaluative goal ambiguity, measured by the number of evaluation criteria that appear towards each goal and the frequency that each individual criterion is used. As can be seen in Table 4, the number of evaluation criteria used to measure program performance towards meeting goals varies greatly when considering both the goal domain and the agency levels. Figure 6 displays the number of evaluation criteria discovered and the frequency with which they were found in the RFP documents analyzed. The ordering of evaluation criteria in Figure 6 corresponds to the order in which they are shown in Table 4.

Figure 6 shows that the number of evaluation criteria used to measure success against deployment goals ranges from 1 to 19, dependent on the goal domain. In addition, Figure 6 (and Table 4) suggests that environmental goals, and reductions in National Ambient Air Quality

Standard (NAAQS) emissions in particular, were the only evaluation criteria used in more than 50 percent of the programs analyzed.

***Figure 6**: Representation and Number of Evaluation Criteria by Goal Domain*



Five (5) environmental evaluation criteria were identified with reductions in NAAQS emissions representing the most frequently used evaluation criteria. The low number of evaluation criteria and the similarity in how often specific evaluation criteria are used, indicates that there is a low level of evaluative goal ambiguity for environmental goals. Notably, reductions in GHG emissions were found in much fewer of the procurement documents analyzed. This finding is perhaps indicative of the period of analysis (2006-2011) in which GHG emissions had not until this point been a requirement of many policies (see Gallagher et al., 2007 for a detailed discussion on this topic). As this requirement becomes more prevalent in environmental policy it is likely that reductions in GHG emissions will become a more prevalent evaluation criterion. Similarly, as policy begins to incorporate considerations governing water pollution rates, land use impacts, and competing fuels and fuel production methods (e.g. the “well to wheels” analysis method and similar - see GMC et al., 2001; Brinkman et al., 2005; and Wu et al., 2006), these factors are likely to appear more frequently as evaluation criteria as well.

The number of evaluation criteria used to measure project and program effectiveness against economic goals is similar in number to those used to measure environmental performance. However, these evaluation criteria are not as well represented across programs in comparison. Because these evaluation criteria are less represented and appear with approximately the same frequency, there does not appear to be a clear favorite in terms of how agencies should measure progress against goals relating to economic growth. Consequently, there is more evaluative goal ambiguity for measuring progress towards economic goals than with environmental goals.

As was previously discussed, gallons of petroleum displaced represented the only evaluation criterion directly related to the goal of enhancing energy security. This evaluation criterion was found in 36 percent of the procurement documents analyzed. Thus, there appears to be some degree of consensus in terms of how to measure progress toward energy security goals. Consequently, there is an extremely low level of evaluative goal ambiguity surrounding energy security given the propensity for agencies to use a single evaluation criterion in this domain. However, these conclusions rely on the categorization system developed by the authors. As was previously discussed, there is an overarching incongruity in how energy security is defined and ultimately what criteria are used to measure progress against the goal of energy security. As with priority goal ambiguity these findings are dependent on how energy security is defined. Definitions that do not distinguish between environmental, economic, technology transition, and energy security goal domains as we have will result in different findings for this goal domain.

Conversely, there are numerous evaluation criteria associated with the technology transition goal domain. Figure 6 shows that agencies employed 19 unique criteria for evaluating progress against the goal of technology transition. The number of AFVs purchased or subsidized and the vehicle-miles traveled (VMT) with an AFV were some of the most commonly used evaluation criteria. In addition, the number of outreach or education events and the contribution to future market strength were also commonly used. However, these still appeared in approximately a third of the procurement documents analyzed. Consequently, it is fair to say that the highest levels of evaluative goal ambiguity are found in the evaluation of progress towards technology transition goals.

INTERVIEWS WITH AGENCY OFFICIALS

Semi-structured interviews were also conducted as part of the research program to serve as both convergent validity for the findings of the content analysis and to gain greater insight into "how" and "why" these findings occur and how improvements might be achieved. Interviews were conducted at the state and regional agency levels. While attempts were made to acquire an interview with an appropriate federal agency level official to discuss the findings (e.g. the Department of Energy, the Environmental Protection Agency, or similar), our interview requests were unsuccessful and are therefore not included. The semi-structured interviews were conducted with officials with direct experience and input in the decision making process for defining and maintaining data for performance measures used in their respective AF deployment grant programs. Three interviews were conducted in total with one interview participant representing a regional agency with a robust AF deployment program, one currently representing a regional agency but also with state energy office experience in similar AF deployment programs, and one representing a highly developed state energy program in AF deployment.

The first series of questions pertained to the overarching concept of benefit from programmatic collaboration between AF deployment programs at federal, state, and regional levels and the current level of collaboration. When asked about the potential of inter-agency collaboration, all interview participants agreed that a greater degree of goal and evaluation alignment between agencies could have a significant and positive impact on the effectiveness of AF deployment efforts. One regional agency interview participant stated that "…right now, agencies are not looking at what other agencies are doing". The interview respondent went further in stating that "none of the metrics overlap, whatsoever" and that "we are not promoting collaboration". The state level interview participant stated "the DOE does so much great work and it is a lost opportunity that there isn't more collaboration with state agencies". Thus the consensus of the interview participants was that greater goal and evaluation alignment can produce greater synergy beyond the level that is currently realized.

The second series of questions pertained to "why" there is goal and evaluation misalignment between agencies. One of the regional agency interview participants stated that "agencies only get credit towards goals that were defined for those agencies". "Agencies are tasked with different goals by design…therefore they only get credit towards these goals and no other credit towards anything that is outside of those goals". "This can result in deployment programs that conflict, sometimes even within the same agency". The state agency respondent also addressed misalignment caused by a top-down approach. This respondent stated that "when the DOE writes a solicitation, they are locked in to a schedule and an approach". At that point, the state agency becomes "…just another stakeholder" and lacks the ability to give or receive input to better align the goals and evaluation criteria of the programs. To provide an example of this, the interview participant spoke of the AF deployment grants funded under the American Recovery and Reinvestment Act (ARRA). The state agency respondent described this series of grants as a “missed opportunity” since state energy agencies were not partners in the grant writing process. Rather, they were tasked with simply dispersing monies allocated to them to serve predefined tasks. While the intentions were good, many of the AF deployment programs funded were misaligned with the goals of the state agency, and therefore missed opportunities for greater synergy towards common goals.

The third series of questions in the interviews pertained to “where” and “when” greater goal and evaluation alignment could be achieved between agencies and programs. All interview participants were in consensus in stating that greater collaboration is most easily achieved between state and regional agencies. The regional level interview participants stated that “…there is often political will and continuity at state and regional levels making collaboration more likely”. In contrast, the interview participants stated that political will and continuity from the federal level is commonly variable through time as administrations, and therefore goals, change. A regional agency official stated that from the federal level, “money and time needs to be put into this continually and consistently”. Consistency in purpose and funding from the federal level will help state and regional agencies maintain consistency as well. This phenomenon is well supported by the literature in this area (see Sperling, 1988; Gallagher, et al., 2007; Sperling and Gordon, 2009). In reference to “when” collaboration should take place, the state agency interview participant stated that “…it is too late to collaborate when you get to the grant writing stage”. Collaboration towards goal and evaluation alignment therefore must occur upstream of this process. The regional agency interview participant also stated that “collaboration has to happen before money is allocated to an agency to achieve a goal”. Therefore, it appears that collaboration should occur in the developmental stages of deployment programs. The greater the development of a program, the harder it becomes to align goals and evaluation criteria.

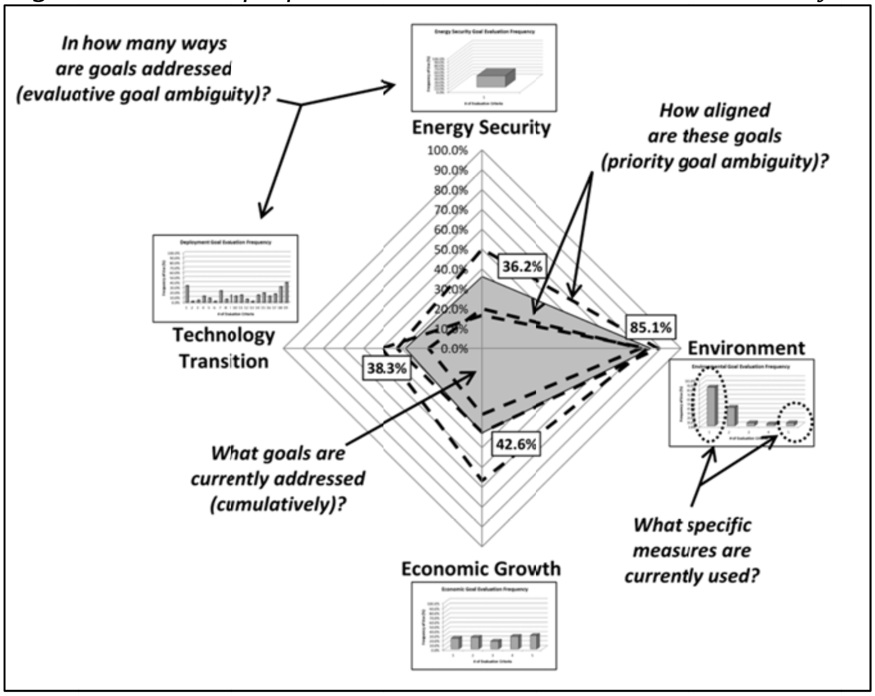
The fourth, and perhaps most important, series of questions in the interviews pertained to “how” greater collaboration can be achieved between agencies and programs. The first recommendation asserted by a regional agency interview participant was to “find common metrics that credit agencies for using common goals and common metrics”. Stated simply, agencies must receive credit for collaboration in order for collaboration to become a goal. Including performance measures that evaluate inter-agency and inter-program collaboration is therefore a primary recommendation for “how” to better implement a more cooperative system. The state level agency interview participant stated that there needs to be a more distinct differentiation between short and long term benefits and goals of deployment programs. They cited natural gas vehicles in comparison to hydrogen or electric drive vehicles as a primary example. They stated that “it is very easy to calculate a cost/benefit ratio for natural gas vehicles because the benefits are immediate”. In contrast calculating a cost/benefit ratio for hydrogen or electric drive vehicles is far more complex due to the fact that the technology is less proven and less deployed. However, the latter may have greater long term benefits when greater market penetration is achieved. Therefore, programs should distinguish between short and long term goals of AF deployment. In addition, more specific and common performance measures for long term deployment benefit need to be defined and implemented.

The fifth, and final, line of questions pertained to the problems associated with the energy security goal domain specifically. As was previously discussed, there are many disparate ways in which contributions towards energy security goals can be (and are) directly measured. In addition, there is a great deal of variability in the definition of energy security. What does energy security mean when considering the transportation sector and AFs specifically? The most common way to measure this is by using the metric of gallons of gasoline or diesel displaced by the use of AFs or AFVs. While this has been a metric commonly associated with the use of biofuels specifically, the regional agency level interview participants stated that this metric is commonly used for measuring the effectiveness of all alternative fuels. One regional level interview participant stated that “gallons of gasoline displaced is the most direct way to measure this” and “…having one measure that is mutually understood across all agencies makes the use of this metric really easy and useful”. However, the state level interview participant stated that “we never correlate this back to the more predominant factors of energy security such as the increase or decrease in foreign oil imported – there are too many moving parts”. Therefore a primary means for “how” to improve alignment within this goal domain is to provide a more common definition for energy security, while also defining what measures best correlate to progress within this goal domain. There is a distinct need for greater research in this area.

APPLICATIONS AND GOAL MAPPING

In this paper two primary measures of goal ambiguity are described: priority goal ambiguity and evaluative goal ambiguity. In addressing these areas, we present a high-level analysis of the goals currently addressed and how performance towards those goals is currently assessed. Figure 7 depicts the areas of information addressed by this research and the combined results.

Figure 7: Areas of Information Addressed and Combined Analysis Results



However, an equally important application of the methods presented in this paper is mapping and tracking goal ambiguity going forward. Such a process is imperative for iteratively assessing changing levels of goal ambiguity over time and adjusting AF deployment policies accordingly to increase the efficiency and effectiveness of these policies and policy mechanisms. Tracking priority goal ambiguity provides a knowledge base of what goals are addressed by AF deployment efforts and by whom. Tracking evaluative goal ambiguity provides a means for assessing the commonality of definitions and metrics used to assess the performance of AF deployment efforts.

As was presented previously, different levels of priority goal ambiguity may be viewed positively or negatively. Centralized versus non-centralized approaches to goals may or may not be more efficient and effective. In the context of this research, a centralized approach is synonymous with low levels of priority goal ambiguity and a decentralized approach is synonymous with high levels of priority goal ambiguity. While analyzing the relative efficiency of decentralized versus centralized approaches is beyond the scope of this research, mapping the level of priority goal ambiguity over time provides a foundation for comparing the level of ambiguity (i.e. level of centrality) with perhaps an optimized level or at least with multiple options that are based on research. The body of literature on environmental federalism, polycentrism, and top-down/bottom-up approaches (see Banzhaf and Chupp; Sovacool, 2011; and Lutsey and Sperling, 2008 for examples) may provide the foundation for similar research-based comparisons. Moreover, tracking this relationship through time provides policymakers and agencies with the information necessary to iteratively improve AF deployment policies and to understand what approaches have or have not improved deployment efforts. Tracking evaluative goal alignment over time is of equal importance. As was presented previously, there is a need for common metrics and common information about progress towards AF deployment goals. This research presents a means for tracking changes in evaluative goal alignment over time.

While the research presented in this paper are only indicative of a narrow band of AF deployment mechanisms (grants in this case), this methodology can be applied to any of the variety of deployment mechanisms currently in use. Doing so would enable agencies at all levels of government to benchmark current alignment and to devise plans for improving programmatic goal collaboration and metric alignment over time. The process of goal alignment mapping can also be applied to other phases of technology diffusion (e.g. technical research and development). In doing so, we recommend an analysis of evaluated goals over stated goals. As was previously stated, analysis of evaluated goals over stated goals is a better reflection of reality due to the common practice of not directing funds towards stated goals within technology deployment programs (IEA, 2002). The analysis of documents conducted as part of this research uncovered several examples of this phenomenon. The use of only evaluative goals for the analysis of goal alignment, as was demonstrated in this paper, can enable agencies to eradicate such contradictions.

CONCLUSIONS

This paper presents a methodology for describing and assessing the alignment of goals for deploying AFs, and demonstrates the use of this methodology in assessing the current state of practice for fuel-neutral and project-neutral (wide-scope) grant programs in the United States. The research revealed that wide-scope AF deployment grant programs most commonly target environmental goals. It further finds that agencies are highly aligned in doing so, both in terms of *what* specific goals are addressed and *how* progress is evaluated. Conversely, economic, energy security, and deployment goals were less represented in AF deployment grant programs and with higher degrees of priority goal ambiguity. In addition, the number of ways used to evaluate "success" against these goals (evaluative goal ambiguity) was highest in the technology transition goal domain.

Environmental goals were most commonly and consistently addressed by federal, regional, and state agencies, resulting in low levels of priority goal ambiguity. Additionally, agencies commonly specified NAAQS emission reduction metrics for quantifying success against environmental goals. Consequently, such programs revealed a relatively low level of evaluative goal ambiguity as well. It is worth noting, however, that the environmental goals of such programs rarely targeted GHG emissions, life cycle water pollution, or life cycle water/land usage. Economic goals were found to be addressed most commonly by federal agencies, and less so by state and regional agencies that work to deploy AFs through wide-scope grant programs. While the research uncovered only five unique categories of evaluation criteria related to economic goals, the dispersion in use of these measures finds that consensus is lacking over the appropriate evaluation criteria for gauging success against economic objectives.

Interestingly, the research finds that state grant programs most often measure energy security goals, while federal and regional level grant programs do so to a lesser degree. Consequently, priority goal ambiguity for this goal domain is relatively high, suggesting much lower alignment. Agencies commonly use a single evaluation criterion when measuring against energy security objectives, which suggests an extremely low level of evaluative goal ambiguity. However, there are many varying interpretations of the term “energy security” therefore we present these results with the caveats presented in the “Limitations” section of this paper. Deployment goals were found to have a relatively high level of priority goal ambiguity, with federal and state programs preferring to cite this goal more often than regional programs. Agencies articulated deployment goals with 19 unique evaluation criteria, resulting in the highest level of evaluative goal ambiguity of any of the goal domains studied.

This research also demonstrates how alignment within other technology deployment programs at various levels of government (e.g. tax incentives, rebates, guaranteed loans, and similar), as well as other research and development programs for AFs, could be evaluated by the methodologies discussed here. Doing so would enable agencies to objectively assess the current status of programmatic alignment and to target those areas most in need of improvement.

LIMITATIONS AND FUTURE RESEARCH

While this research identifies the current level of agency integration in supporting AFs towards multiple objectives, it cannot exhaustively explain *why* agencies lack integration in these areas. While some insights were discovered in interviews with agency officials, this area of research should be explored further and will likely have to implement a broad, multi-disciplinary (and perhaps time-series) approach for doing so. In addition, we cannot state what the exact effect of developing and implementing more integrated policy mechanisms will be. Literature consistently states that better integrated policy will result in more effective results (in this case more effective deployment). However, there is little knowledge on what the magnitude of those effects will be. As a result, there is a need for research in this area so that these effects may be better described.

In addition, there is a quandary regarding the superiority of a centralized policy approach (i.e. less goal ambiguity) against a decentralized approach (i.e. greater goal ambiguity). Research suggests that there are instances in which one approach may be superior over the other in terms of the efficiency and effectiveness of policy outcomes. This research falls under many headings including fiscal federalism, environmental federalism, polycentrism, and top-down/bottom-up approaches (to name a few). Research in the specific context of AF deployment and within the goal domains outlined in this research may show that specific goal domains may be addressed more efficiently through more or less centralized approaches. As a result, there is a distinct need for research in this area, specifically in the context of AF deployment efforts. Results from this type of research could be used to compare progress against the current state of centrality by using the categorization and analysis methods presented in this paper.

In this research we have attempted to isolate energy security as a goal separate from environmental, economic, and technology transition goals. However, there are varying interpretations of the term “energy security” and other research on energy security use definitions that overlap the goal domains presented in this paper. Therefore, we present these results with the caveat that different interpretations of energy security will result in different research outcomes than are presented in this paper.

Another area of interest is the reliability of the evaluation criteria themselves. While the presence of the evaluation criteria are identified in this paper and used for analysis, many of the definitions for these evaluation criteria differ in terms of the time period that they apply to (e.g. the period for which AFVs will be in service) and the clarity of the definition of the criterion itself (e.g. does a job created mean the same thing to all agencies and grant proposers). Analysis of this topic might provide an additional dimension of knowledge for goal alignment in future studies.

Of additional interest are the relationships between meeting each of these goals individually. For example, what effect does an increase in jobs creation have on the long-term effectiveness of meeting environmental goals? As is suggested by Churchman, the objective of a system can be measured by how much “…the system will knowingly sacrifice other goals in order to obtain the objective” (Churchman, 1968, p.31). Thus if the true objective is environmental impact, goals such as jobs creation may hinder the AF system from achieving the original environmental objective. By how much is a question for additional research. The answers could vastly improve the design of policy mechanisms for the deployment of AFs.

Finally, it was discovered that agencies most commonly choose to fund both AFs and AFV purchases simultaneously under the same grant, such that refueling infrastructure projects compete against each other, as well as with AFV purchase projects. We state only that this is the current state of practice and can make no assessment of the effectiveness of this approach. Research that can assess the practicality and effectiveness of this approach versus other approaches could prove to be quite useful when considering how to better design these programs.

GRANT PROGRAMS ANALYZED BY ALGENCY LEVEL CATEGORY

Federal Grant Programs

1. Environmental Protection Agency – National Clean Diesel Funding Assistance Program
2. Federal Transit Authority – State of Good Repair Bus and Bus Facilities
3. Department of Energy – Clean Cities Petroleum Reduction Technologies Projects for the Transportation Sector
4. Federal Transit Authority – Clean Fuels Grant & Discretionary Bus and Bus Facilities Programs
5. Environmental Protection Agency – Clean School Bus USA
6. Federal Transit Authority – Funding for Transit Investments for Greenhouse Gas and Energy Reductions Grant

State Grant Programs

1. California Energy Commission – Alternative and Renewable Fuel & Vehicle Technology Program
2. New Hampshire Department of Environmental Services – Alternative Fuel Vehicles and Fueling Infrastructure Program
3. California Energy Commission – Medium and Heavy Duty Advanced Vehicle Technology Program
4. California Reformulated Gas Settlement Fund
5. North Carolina Solar Center – Clean Fuel Advanced Technology Project
6. New York State Energy Research and Development Authority – Advanced Transportation Technologies Program
7. Idaho Transportation Department – Congestion Mitigation and Air Quality Improvement Program

8. Texas State Energy Conservation Office – Transportation Energy Efficiency Alternative Fuels and Technology Stimulus Grant Program
9. Texas Commission on Environmental Quality – Texas Clean Fleet Program
10. Texas State Energy Conservation Office – Alternative Fuel and Hybrid Vehicle Grants
11. Texas Commission on Environmental Quality – Emission Reduction Incentive Grants
12. Maryland Energy Administration – Transportation Grant Program
13. Connecticut Department of Transportation – Connecticut Clean Fuel Program
14. Oklahoma State Energy Office – State Energy Program Formula Grant
15. New York State Energy Research and Development Authority – Low Carbon Transportation Alternatives Program
16. Ohio Department of Development – Ohio Diesel Emissions Reduction Grant
17. New Hampshire Department of Environmental Services – Diesel Emissions Reduction Grant Program
18. South Carolina Energy Office – South Carolina Energy Efficiency Block Grant
19. North Carolina Energy Office – Alternative Fuel and Advanced Vehicle Technology Program
20. Wisconsin Office of Energy Independence – Wisconsin Clean Transportation Program: Alternative Fuel Vehicles and Infrastructure Technical Assistance Program
21. Pennsylvania Department of Environmental Protection – Alternative Fuels Incentive Grant
22. California Environmental Protection Agency – Goods Movement Emissions Reduction Program
23. New York State Energy Research and Development Authority – New York State Clean Fueled Bus Program
24. New York State Energy Research and Development Authority – New York State Clean Cities Challenge Program
25. Utah Department of Environmental Quality - Clean Fuel Vehicle Technology Grant
26. North Carolina Department of Environment and Natural Resources – Diesel Emission Reduction Grant

Regional Grant Programs

1. South Coast Air Quality Management District – Alternative Fuel Infrastructure Funding Opportunities For New, Expanded, & Upgraded Refueling Facilities in the South Coast Air Quality Management District
2. Association of Central Oklahoma Governments – Public Fleet Conversion Grants
3. Indian Nations Council of Governments/Tulsa Area Clean Cities Coalition – Transportation Technologies: Public Fleet Conversion Program
4. South Coast Air Quality Management District – Clean Fuels Program
5. North Texas Council of Governments – Clean Fleets of North Texas

6. San Joaquin Valley Air Pollution Control District – Emission Reduction Technology Advancement Program
7. Kern County Air Pollution Mitigation Fund
8. Houston-Galveston Area Council – Clean School Bus Houston
9. Illinois Association of Regional Councils – Energy Efficiency and Conservation Block Grant
10. Appalachian Regional Council – Planning and Implementation of Community Based Energy Projects
11. Maine, New Hampshire, and Vermont Clean Cities – Alternative Fuel Infrastructure in Northern New England
12. Indianapolis Metropolitan Planning Organization – Congestion Mitigation and Air Quality Improvement Program
13. Southeast Michigan Council of Governments – Congestion Mitigation and Air Quality Improvement Program
14. Mid-America Regional Council – Kansas City Metropolitan Congestion Mitigation/Air Quality Fund
15. Northern Sierra Air Quality Management District – California DMV Surcharge Fund Program

CHAPTER REFERENCES

Alternative Fuels Data Center (AFDC), 2010. Data, Analysis, and Trends. United States Department of Energy, Alternative Fuels Data Center (online), retrieved February 17, 2010 from http://www.afdc.energy.gov/afdc/data/vehicles.html.

Alternative Fuels Data Center (AFDC), 2011. Federal & State Incentives & Laws. United States Department of Energy, Alternative Fuels Data Center (online), retrieved July 3, 2011 from http://www.afdc.energy.gov/afdc/laws/.

Bang, G., 2010. Energy Security and Climate Change Concerns: Triggers for Energy Policy Change in the United States? Energy Policy, 38(4), 1645-1653.

Banzhaf, H.S., Chupp, B.A., 2010. Heterogeneous Harm vs. Spatial Spillovers: Environmental Federalism and US Air Pollution. National Bureau of Economic Research (NBER), Working Paper Series, NBER Working Paper 15666.

Brinkman, N., Wang, M., Weber, T., Darlington, T., 2005. Well-to-Wheels Analysis of Advanced Fuel/Vehicle Systems — A North American Study of Energy Use, Greenhouse Gas Emissions, and Criteria Pollutant Emissions, General Motors Corporation, Argonne National Laboratories, and Air Improvement Resource, Inc., Phase 2 Report May 2005.

Bureau of Transportation Statistics (BTS), 2011. National Transportation Statistics 2011, United States Department of Transportation, BTS (online), retrieved July 7, 2011 from http://www.bts.gov/publications/national_transportation_statistics/pdf/entire.pdf.

Byrne, M.R., Polonsky, M.J., 2001. Impediments to Consumer Adoption of Sustainable Transportation – Alternative Fuel Vehicles. International Journal of Operations and Production Management, 21(12), 1521-1538.

Calciolari, S., Cantu, E., Fattore, G., 2011. Performance Management and Goal Ambiguity: Managerial Implications in a Single Payer System. Health Care Management Review, 36(2), 164-174.

Chester, L., 2010. Conceptualising Energy Security and Making Explicit its Polysemic Nature. Energy Policy, 38(2), 887-895.

Churchman, C.W., 1968. The Systems Approach, Dell Publishing Co., New York, New York.

Chun, Y.H., Rainey, H.G., 2005. Goal Ambiguity and Organizational Performance in U.S. Federal Agencies. Journal of Public Administration Research Theory, 15(1), 1-30.

Deutch, J.M., 2011. An Energy Technology Corporation Will Improve the Federal Government's Efforts to Accelerate Energy Innovation, Brookings Institution, Discussion Paper 2011-05, May 2011.

Gallagher, K.S., Collantes, G., Holdren, J.P., Lee, H., Frosch, R., 2007. Policy Options for Reducing Oil Consumption and Greenhouse-Gas Emissions from the U.S. Transportation Sector, Energy Technology Innovation Policy, Discussion Paper, July 2007.

General Motors Corporation (GMC), Argonne National Laboratory (ANL), BP, ExxonMobil, and Shell, 2001. Well-to-Wheel Energy Use and Greenhouse Gas Emissions of Advanced Fuel/Vehicle Systems — North American Analysis, Volumes 1-3, available online at http://greet.es.anl.gov/publications.

Greene, D.L., 2010. Measuring Energy Security: Can the United States Achieve Oil Independence? Energy Policy, 38(4), 1614-1621.

Holsti, O.R., 1969. Content Analysis for the Social Sciences and Humanities, Addison-Wesley Publishing, Reading, Massachusetts.

Howell, D., Chelius, M., 1997. Alternative Fuel Vehicle Programs – Applicability of Government Incentives. Transportation Research Record, Volume 1587, 1-9.

International Energy Agency (IEA), 2002. Deployment Strategies for Hybrid, Electric, and Alternative Fuel Vehicles, Annex VIII Final Report.

International Energy Agency (IEA), 2011. Technology Roadmap – Electric and Plug-In Hybrid Electric Vehicles, Updated June 2011.

Jansen, J.C., Seebregts, A.J., 2010. Long Term Energy Services Security: What is it and how can it be measured and valued. Energy Policy, 38(4), 1654-1664.

Jung, C.S., 2011. Developing and Validating New Concepts and Measures of Program Goal Ambiguity in the U.S. Federal Government. Administration and Society, (online before print), 1-27.

Kruyt, B., van Vuuren, D.P., de Vries, H.J.M., Groenenberg, H., 2009. Indicators for Energy Security. Energy Policy, 37 (6), 2166–2181.

Latham, G.P., Locke, E.A., 1991. Self-regulation through goal setting. Organizational Behavior and Human Decision Processes, 50, 212-247.

Lee, H., 2009. Oil Security and the Transportation Sector, in: Gallagher, K.S. (Ed.), Acting in Time on Energy Policy. Brookings Institution Press, Washington, D.C.

Locke, E.A., 1968. Toward A Theory of Task Motivation and Incentives. Organizational Behavior and Human Performance, 3, 157-189.

Loschel, A., Moslener, U., Rubbelke, D., 2010. Indicators of Energy Security on Industrialised Countries. Energy Policy, 38(4), 1665-1671.

Lutsey, N., Sperling, D., 2008. America's Bottom-Up Climate Change Mitigation Policy. Energy Policy, 36(2), 673-685.

Matland, R. E., 1995. Synthesizing the Implementation Literature: The Ambiguity-Conflict Model of Policy Implementation. Journal of Public Administration Research and Theory, 5, 145-174.

May, P., Saptotichne, J., Workman, S., 2006. Policy Coherence and Policy Domains. Policy Studies Journal, 34(3), 381-403.

Mazmanian, D., Sabatier, P., 1983. Implementation and Public Policy, Scott Foresman, Glenview, Illinois.

Melaina, M., McQueen, S., Brinch, J., 2008. Refueling Infrastructure for Alternative Fuel Vehicles: Lessons Learned for Hydrogen – Workshop Proceedings, National Renewable Energy Laboratory (NREL), Proceedings NREL/BK-560-43669, July 2008.

Melendez, M., 2006. Transitioning to a Hydrogen Future: Learning from the Alternative Fuels Experience, National Renewable Energy Laboratory (NREL), Technical Report NREL/TP-540-39423, February 2006.

National Renewable Energy Laboratory (NREL), 2011. Commercialization and Deployment at NREL – Advancing Renewable Energy and Energy Efficiency at Speed and Scale, NREL/MP-6A42-51947, May 2011.

Neuendorf, K.A., 2002. The Content Analysis Guidebook, Sage Publications, Thousand Oaks, California.

Pandey, S.K., Wright, B.E., 2006. Connecting the Dots in Public Management: Political Environment, Organizational Goal Ambiguity, and the Public Manager's Role Ambiguity. Journal of Public Administration Research Theory, 16(4), 511-532.

Ripley, R.B., Franklin, G.A., 1982. Bureaucracy and Policy Implementation. The Dorsey Press, Homewood, Illinois.

Sagar, A., Gallagher, K.S., 2004. Energy Technology Demonstration and Deployment, in: National Commission on Energy Policy (NCEP) Technical Appendix for Chapter 6: Developing Better Energy Technologies for the Future, in: Ending the Energy Stalemate – A Bipartisan Strategy to Meet America's Energy Challenges, December 2004.

Sharma, C.H., 2011. Beyond Gaps and Imbalances: Restructuring the Debate on Intergovernmental Fiscal Relations. Public Administration, 90(1), 99-128.

Shen, Y., Lin, G., Li, K., Yuan, B., 2010. An Assessment of Exploitable Renewable Energy Sources with Concerns of Policy and Technology. Energy Policy, 38(8), 4604-4616.

Sovacool, B.K., 2011. An International Comparison of Four Polycentric Approaches to Climate and Energy Governance. Energy Policy, 39(6), 3832-3844.

Sperling, D., Gordon, D., 2009. Two Billion Cars: Driving Towards Sustainability, Oxford University Press, New York, New York.

Sperling, D., Yeh, S., 2009. Low Carbon Fuel Standards. Issues in Science and Technology (online), Winter 2009, retrieved June 16, 2011 from http://www.issues.org/25.2/sperling.html.

Sperling, D., 1988. New Transportation Fuels – A Strategic Approach to Technological Change, University of California Press, Berkeley, California.

United States Congress (USC), 1992. Energy Policy Act of 1992, Public Law 102-486, 102nd Congress.

United States Congress (USC), 2009. American Recovery and Reinvestment Act, Public Law 111-5, 111th Congress.

United States Department of Energy (USDOE), 2011. Database of State Incentives for Renewables and Efficiency (DSIRE). United States Department of Energy, Energy Efficiency and Renewable Energy (online), retrieved July 3, 2011 from http://www.dsireusa.org/.

United States Energy Information Administration (USEIA), 2011. Alternatives to Traditional Transportation Fuels 2009, Released April 2011, USEIA (online), retrieved July 7, 2011 from http://www.eia.gov/renewable/alternative_transport_vehicles/pdf/afv-atf2009.pdf.

United States Government Accountability Office (USGAO), 2011. Opportunities to Reduce Potential Duplication in Government Programs, Save Tax Dollars, and Enhance Revenue. Report GAO 11-318SP, March 2011.

Vogel, D., Toffel, M., Post, D., Aragon, N., 2010. Environmental Federalism in the European Union and the United States. Harvard Business School (HBS), Working Paper 10-085.

Winzer, C., 2012. Conceptualizing Energy Security. Energy Policy, Vol. 46, 36-48.

Wu, Y., Wang, M., Sharer, P., Rousseau, A., 2006. Well-to-Wheels Results of Energy Use, Greenhouse Gas Emissions, and Criteria Air Pollutant Emissions of Selected Vehicle/Fuel Systems. SAE 2006 Transactions (Journal of Engines), Paper No. 2006-01-0377.

CHAPTER 3: METRIC UNCERTAINTY WITHIN SELECTION AND EVALUATION CRITERIA OF ALTERNATIVE FUEL DEPLOYMENT PROJECTS

ABSTRACT

Transitioning to fuel sources that are alternative to traditional gasoline and diesel platforms within the on-road transportation sector is a common goal of agencies. Measurement of performance towards these goals is less clear, especially when considering the goal domains of technology transition and energy security. In this paper, a survey methodology is used to characterize attributes related to the reliability of common metrics used to evaluate project and program performance towards these goals in alternative fuel deployment grant programs. Survey responses were accumulated from 46 federal, state, and regional agency officials with direct experience in reporting performance on alternative fuel deployment grant programs. The results from the survey show that units of fuel displacement, the number of alternatively fueled vehicles distributed, and the number of alternative fuel refueling stations installed are the most reliable metrics currently incorporated within deployment programs. Conversely, the direct measurement of vehicle miles traveled (VMT) and customer use of refueling stations presents measurement challenges in terms of the feasibility of acquiring the necessary data, the objectivity with which data are reported, and the clarity of definitions used for these metrics. Measuring performance in terms of refueling station density and reductions in corridor refueling gap distance is shown to have great potential but is currently underutilized. The measurement of performance in outreach and education efforts was found to have the highest levels of metric uncertainty and constitutes a major need for future research efforts. Systematic evaluation and improvement within measurement systems can help guide practitioners and policymakers in how to better design evaluation systems in the future.

1. INTRODUCTION

The deployment of alternative fuels (AFs) for the on-road transportation sector has and continues to be an objective of different agencies and programs. Yet as agencies attempt to meet the many goals that AFs are designed to address, there is little consensus on which goals to address and how to systematically measure progress, especially in the areas of energy security and technology transition (Sobin, et al., 2012). While the deployment of AF technologies have been somewhat successful in addressing environmental goals, alternative fuel vehicles (AFVs) currently account for less than 2 percent of licensed on-road vehicles in the United States (US) and less than 1 percent of use and throughput in comparison to gasoline and diesel fuels (USEIA, 2011; BTS, 2011). We surmise that this is, at least in part, a consequence of metric uncertainty within the measurement of progress towards technology transition and energy security goal domains.

The alignment of goals and evaluation systems towards those goals in the context of AF deployment programs have been only peripherally understood and studied to date. While numerous texts have alluded to the need for alignment and consistency within AF deployment strategies, policies, and evaluation systems (United States Government Accountability Office (USGAO), 2011; Deutch, 2011; International Energy Agency (IEA), 2011; National Renewable Energy Laboratory (NREL), 2011; Sperling and Yeh, 2010; Sperling and Gordon, 2009; Lee, 2009; Melaina, et al., 2008; Melendez, 2006; Sagar and Gallagher,2004; International Energy Agency(IEA), 2002; Byrne and Polonsky, 2001; Howell and Chelius, 1997; Sperling, 1988), few studies have looked directly at the selection and evaluation of metrics used to evaluate performance towards goals.

The term metric uncertainty can take on several definitions as the term can refer to many types of ambiguity or uncertainty in program and policy analysis, as well as numerous other scientific and quality science areas of study. In this paper, we define metric uncertainty as simply the degree to which specific performance metrics are able to be consistently used and interpreted between programs in the broader context of AF deployment programs. In terms of policy analysis literature, we can think of metric uncertainty as an extension of the concept of evaluative goal ambiguity defined as "*the degree of difficulty in objectively evaluating progress toward the achievement of organizational goals*" (Chun and Rainey, 2005, p.534). Decreases in goal ambiguity have been consistently shown to increase progress towards goals in many areas including the healthcare industry, human service agencies, and federal agency programs (Calciolari, et. al., 2011; Pandey and Wright, 2006; Chun and Rainey, 2005). Therefore the purpose of this paper is to further contribute to the methodological literature of studying evaluative goal ambiguity in the context of AF deployment programs in the United States. Ultimately, the goal is to highlight areas of evaluative ambiguity within AF deployment programs and improve them, thereby increasing the rate of deployment for AF technologies in the on-road transportation sector.

Uncertainty associated with metrics related to the deployment of alternative fuels and alternative energy sources has been performed in several previous studies. However in almost all cases, metric uncertainty is described in relationship to environmental goals (see Parkinson et al., 2001; Gupta, et al., 2003; NAS, 2005; Nahorski and Horabik, 2008, Plevin, 2010 for example). While the importance of addressing environmental objectives cannot be overstated, the goals within this domain are well aligned (Sobin, et al., 2012) and well-studied as life-cycle analysis methods related to overall environmental impact continue to develop. Conversely, goal domains of energy security and technology transition are much less aligned (Sobin, et. al., 2012) and are therefore much less understood. Therefore this research looks at the metric uncertainty associated with these domains specifically.

As was previously stated, the definition we use for metric uncertainty in this paper is the degree to which specific performance metrics are able to be consistently used and interpreted between programs in the broader context of AF deployment programs. Literature in the areas of performance measurement and in policy analysis routinely state that performance metrics should be consistent, clearly defined, objective, compatible, measurable, equivocal, repeatable, and reproducible, to name a few of the attributes (Hurst, 1980; Smith, 1989; Kaplan and Norton, 1992; Neely, et al., 1995; Heimann, 1995; Kennerley and Neely, 2002; Behn, 2003; NAS, 2005). It is from these sources of literature that the primary attributes of metric uncertainty were compiled from the literature. The following four attributes of performance metrics tested in this study include:

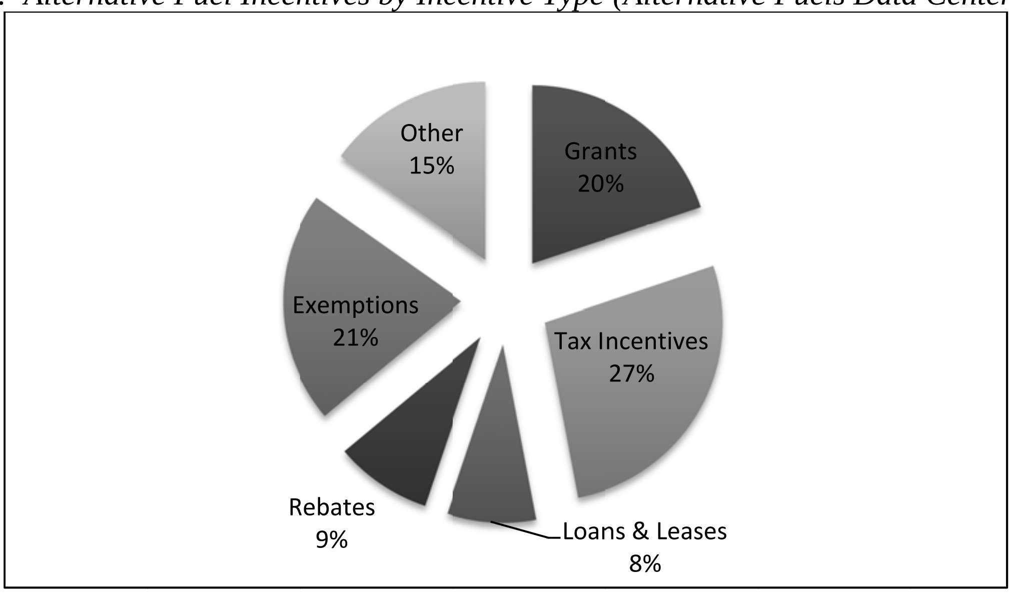
- Appropriateness – the level to which the metric is able to discern differences among competing AF deployment projects;
- Clarity – the level of understanding associated with the definition and boundaries of the measurement unit used in the metric;
- Objectivity – the level of human judgment that must be used when reporting values for each metric; and
- Feasibility – the level of data attainability needed to accurately report values for each metric.

While these attributes should not be considered as a comprehensive list of all positive characteristics of performance metrics, an understanding of how metrics compare within these attributes will begin to bring light to the potential benefits and drawbacks of their use in selecting and evaluating the performance of AF deployment projects.

2. WHERE TO STUDY METRIC UNCERTAINTY IN AF DEPLOYMENT PROGRAMS?

One of the primary difficulties with studying metric uncertainty in the evaluation process is where to focus the effort of research. A holistic view of AF deployment efforts shows that there exists a tangled web of policy mechanisms and incentive types that are used to help deploy on-road AF technologies (Sperling and Gordon, 2009; Sperling, 1988, Howell and Chelius, 1997). To bring light to the current distribution of incentives for AFs by incentive type, Figure 1 presents the current distribution of incentives in the United States (Alternative Fuels Data Center (AFDC), 2013).

***Figure 1**: Alternative Fuel Incentives by Incentive Type (Alternative Fuels Data Center, 2013).*



As is shown in Figure 1, grant programs represent approximately 20 percent of federal and state incentives used to aid the development and deployment of AF technologies. In addition, previous research has shown that many more grant programs exist than are included in this database (see Sobin, et al., 2012). Therefore grants remain one of the most common types of incentives used to deploy AFs.

In addition to the frequency of use, grant programs are also unique in that typically a formal selection and evaluation process is used. That is, grant programs commonly are designed to provide flexibility in choosing among competing fuel technologies (e.g. ethanol versus plug-in electric technologies) and competing project focuses (e.g. purchase of AFVs versus construction of refueling infrastructure or both). This method provides a “wide-scope” approach for selecting the best-value project for a given area that most other incentive types do not. Therefore grant programs are an important, flexible policy tool for deploying AFs. Perhaps more importantly, this flexibility in the project selection and evaluation process means that common metrics must be predefined and used to evaluate the potential and performance of competing projects. It is the transparent and observable nature of these metrics that makes grant programs a unique opportunity for studying how these competing projects are evaluated, and in this paper, the associated uncertainty with these metrics.

3. RESEARCH APPROACH

To achieve the research goal of ascertaining the relative uncertainty among metrics commonly used in AF deployment grant programs, a survey methodology was incorporated. The unit of analysis for the survey was the respondent with direct experience in acquiring and reporting data related to specific metrics in AF deployment grant programs. The total population of grant programs that currently utilize grant programs, and therefore performance metrics, is not certain, but can be estimated. Previous studies by the authors found that a significant population of grant programs existed beyond just those listed in the Alternative Fuels Data Center database on incentives (see Sobin, et al., 2012). Therefore the target population for this study includes agencies that oversee grant programs listed in the AFDC incentive database, as well as other agencies that may not be listed. Grant programs are very commonly used at the state and regional levels by the many Department of Energy – Clean Cities Coalitions (CCCs) and by State Energy Offices (SEOs). Clean Cities Coalitions are a series of regional stakeholder groups overseen by the United States Department of Energy and tasked with deploying AF technologies. State Energy Offices are listed in the National Association of State Energy Offices (NASEO) and, perhaps counter intuitively, sometimes reside within state environmental offices or offices of economic development. In each case, a respondent with direct experience in overseeing and reporting data on the AF deployment grant program using metrics was sought.

Selecting a group of representative metrics for the goal domains of technology transition and energy security can be a complicated process as there are many to choose from. The metrics included in this research were compiled from a previous study (see Sobin, et al., 2012) on AF deployment grant and represent some of the most common metrics used to measure performance towards a wide range of objectives. The metrics included in this research are shown in Table 1.

Table 1: *Metrics of Interest within Energy Security and Technology Transition Goal Domains*

Goal Domain	Objective	Metric of Interest
Energy Security	Dependence on traditional energy sources minimized	Gallons of gasoline/diesel displaced
Technology Transition	Distribution of AF technologies maximized	# of AFVs purchased/subsidized
	Usage of AFVs maximized	Vehicle-miles traveled (VMT) with AFVs
	Distribution system of AFs maximized	# of AF refueling stations constructed
		# of customers served by AF refueling stations constructed
		Increase in refueling station density
		Decrease in refueling corridor gap distance
	AF market transformation and acceptance maximized	# of outreach/education events hosted

Completed survey responses were acquired to test metric uncertainty among the metrics shown in Table 1. Responses were gathered from 46 survey participants representing 29% of all Clean Cities Coalitions and 28% of State Energy Offices listed in the NASEO database. However, several SEOs contacted were found to not currently use grant programs as a means for deploying AF technologies or the metrics described. When NASEO agencies and programs that do not use grant programs to deploy AFs, are excluded the response rate increases to 38%. To ensure a requisite data set, responses were also attained from agencies outside of the AFDC database of grant programs, the list of NASEO members, and the list of Clean Cities Coalitions. The other agencies were selected because they sometimes use grant programs to deploy AF technologies as well. Tables 2, 3, 4, and 5 describe the characteristics of the survey respondents.

***Table 2**: Number and Percentage of Survey Respondents by Agency Level*

Agency Level	# of Responses
Federal	1 (2.2%)
State	19 (41.3%)
Regional (within state)	21 (45.7%)
Regional (multi state)	2 (4.3%)
Other	3 (6.5%)
Σ	**46 (100%)**

***Table 3**: Number and Percentage of Survey Respondents by Agency Focus*

Agency Focus	# of Responses
Energy	30 (65.2%)
Environment	2 (4.3%)
Transportation	8 (17.4%)
Other	6 (13.0%)
Σ	**46 (100%)**

***Table 4**: Number and Percentage of Survey Respondents by Years of Experience*

Years of Experience	# of Responses
1 to 2	6 (13.0%)
3 to 5	19 (41.3%)
6 to 10	7 (15.2%)
11 to 15	7 (15.2%)
>15	7 (15.2%)
Σ	**46 (100%)**

***Table 5:** Number and Percentage of Survey Respondents by Average Annual Budget for AF Deployment Projects*

Average Annual Budget (per year)	# of Responses
Not reported	3 (6.5%)
<$50k	5 (10.9%)
$50k to $100k	4 (8.7%)
$100k to $250k	6 (13.0%)
$250k to $500k	7 (15.2%)
$500k to $1M	9 (19.6%)
$1M to $5M	11 (23.9%)
>$5M	1 (2.2%)
Σ	**46 (100%)**

Respondents to the survey self-reported an average experience level of 8.7 years in deploying AF technologies for the on-road transportation sector through grant programs. In addition, 41% of the respondents self-reported having 10 or more years of experience. From a geographical coverage perspective responses were accumulated from agencies within 31 of the 50 states (62% of states represented by an agency).

The 8 metrics of interest identified in Table 1 were tested by one each of the aforementioned attributes of metric uncertainty:

- Appropriateness – the level to which the metric is able to discern differences among competing AF deployment efforts at the project level;

- Clarity – the level of understanding associated with the definition and boundaries of the measurement unit used in the metric;
- Objectivity – the level of human judgment that must be used when reporting values for each metric; and
- Feasibility – the level of data attainability needed to accurately report values for each metric.

Each survey respondent was asked to rate each metric using a 5-point interval level rating scale relating to the level of performance associated with each attribute/metric combination. To measure agreement among the survey respondents, raw agreement was calculated. However, several sources within literature show that inter-rater agreement (IRA) should also be calculated to compensate for respondents having chance agreement on a scale (James, et al., 1984; Lance, et al., 2006; LeBreton and Senter, 2008; Wagner, et al., 2010, Gwet, 2010). Therefore, compensating for chance related agreements presents a higher standard by which agreement can be measured than by simply reporting raw agreement values. Inter-rater agreement is *"…used to address whether scores furnished by judges are interchangeable or equivalent in terms of their absolute value*" (LeBreton and Senter, 2008, p.816). The statistic used to describe IRA is within group reliability, r_{wg}, and ranges from 0 to 1. A value of 0 represents no IRA and a value of 1 represents full IRA. The equation for r_{wg} is shown in Equation 1. Required intermediate calculations are shown in Equations 2 and 3. Equation 1 was originally developed by Finn (1970) and further explained by James, et al. (1984). James, et al. (1984) credit Mood, et al. (1974) for the development of equation 3.

$$r_{wg} = 1 - \frac{\sigma_o^2}{\sigma_u^2} \quad \text{(Equation 1)}$$

where: σ_o^2 = the observed variance among the respondents' ratings (Equation 2)

σ_u^2 = the expected error variance due to chance agreement (Equation 3)

$$\sigma_o^2 = \frac{\sum(\bar{x}-x)^2}{(N-1)} \quad \text{(Equation 2)}$$

$$\sigma_u^2 = \frac{(c^2-1)}{12} \quad \text{(Equation 3)}$$

where: c = the number of categories on the response scale

The most common thresholds for acceptance substantial agreement among rater using the r_{wg} varies in the literature between 0.6 and 0.8 (see Lance, et al., 2006; Wagner, et al., 2010 for a detailed discussion on this topic). Wagner, et al. state that "*…r_{wg} values of 0.8 and above indicate strong agreement, values between 0.7 and 0.8 indicate moderate agreement, values between 0.6 and 0.7 show weak agreement, while values below 0.6 represent unacceptable levels of agreement*" (Wagner, et al., 2010, p. p.595). While there is some inconsistency in the literature surrounding the interpretation of r_{wg} values above 0.6, the literature is consistent in rejecting r_{wg} values less than 0.6 and we do in this study as well. Moreover, it should be noted in any case that the r_{wg} threshold is a "subjective heuristic" used to describe agreement and is therefore open to interpretation by the reader (LeBreton and Senter, 2008; Wagner, et al., 2010).

4. RESULTS

The results of cumulative raw agreement ratings provided by the respondents are presented in Table 6. The ratings shown in Table 6 are abbreviated to show whether respondents rated each attribute/metric combination on the high end of the scale (an indication of high performance within that combination), on the neutral portion of the scale (neither high or low performance), or on the low end of the scale (an indication of low performance within that attribute/metric combination). Notably there are a different number of responses for each metric. This is due to the survey design which only allowed survey respondents with direct past experience in the use of that metric to provide ratings for that metric.

Table 7 shows the within-group agreement coefficients (r_{wg}) for each attribute/metric combination. A "red-yellow-green" stoplight system is used in this table to denote levels of agreement among the respondents. Cells highlighted in green represent r_{wg} agreement values of 0.7 or greater, indicating at least moderate to strong agreement among the respondents according to the interpretation of r_{wg} values in the literature (see Lance, et al., 2006; Wagner, et al., 2010). Cells highlighted in yellow represent r_{wg} agreement values of 0.60-0.69 which indicates a low but acceptable level of agreement for accepting outcomes from the data. Cells highlighted in red indicate lower than acceptable levels of agreement. Five (5) attribute/metric combinations are not shown (blacked out). In these cases hypothesis testing (discussed in detail in the next section) showed the results to be insignificant. Therefore they are not reported.

Table 8 presents the results of hypothesis testing based on the calculated r_{wg} values. The values shown in Table 8 simply show a "Yes" or "No" based on the results from a chi-square test for significance. In relation to the r_{wg} results "*…the null hypothesis tested by chi-square is that there is no agreement among raters in their rating of an item above and beyond what would be expected by chance or random responding*" (Dunlap, et al., 2003). Notably there is some controversy surrounding the application of the chi-square test to the rectangular distribution assumption used in the original r_{wg} equation (see Equations 1 and 3). In response to this dilemma Dunlap, et al. (2003) provide a modified list of acceptance values based on simulations

***Table 6**: Raw Agreement Values for Attribute/Metric Combinations Tested*

PM Attribute	PM	GFD	# of AFVs	VMT	# of Cust.	# of Stations	# of Outreach	Stat. Density	Corridor Gap
	# of Ratings	42	36	32	18	39	37	5	8
Appropriateness	**Low Overall (%)**	2.38%	5.56%	9.38%	11.11%	2.56%	10.81%	0.00%	0.00%
	Neutral Overall (%)	4.76%	16.67%	21.88%	22.22%	23.08%	37.84%	20.00%	12.50%
	High Overall (%)	92.86%	77.78%	68.75%	66.67%	74.36%	51.35%	80.00%	87.50%
Clarity	**Low Overall (%)**	0.00%	0.00%	3.13%	27.78%	0.00%	16.67%	20.00%	25.00%
	Neutral Overall (%)	11.90%	5.56%	21.88%	27.78%	10.26%	11.11%	20.00%	12.50%
	High Overall (%)	88.10%	94.44%	75.00%	44.44%	89.74%	72.22%	60.00%	62.50%
Attainability (Feasibility)	**Low Overall (%)**	4.76%	0.00%	16.13%	55.56%	0.00%	2.70%	0.00%	0.00%
	Neutral Overall (%)	23.81%	2.78%	29.03%	16.67%	2.63%	21.62%	0.00%	0.00%
	High Overall (%)	71.43%	97.22%	54.84%	27.78%	97.37%	75.68%	100.00%	100.00%
Objectivity	**Low Overall (%)**	2.38%	5.56%	9.38%	11.11%	0.00%	11.11%	20.00%	25.00%
	Neutral Overall (%)	23.81%	0.00%	31.25%	38.89%	5.13%	25.00%	0.00%	0.00%
	High Overall (%)	73.81%	94.44%	59.38%	50.00%	94.87%	63.89%	80.00%	75.00%

***Table 7**: Calculated Within-Group Agreement Values (r_{wg}) for Attribute/Metric Combinations Tested*

PM	GFD	# of AFVs	VMT	# of Cust	# of Stations	# of Outreach	Stat. Density	Corridor Gap
# of Ratings	42	36	32	18	39	37	5	8
Appropriateness	0.75	0.61	0.50	0.56	0.67	0.57		0.71
Clarity	0.77	0.82	0.65		0.77	0.41		
Attainability	0.62	0.88	0.54	0.35	0.87	0.56	0.90	0.86
Objectivity	0.64	0.71	0.46	0.69	0.83	0.52		

***Table 8**: Hypothesis Testing for r_{wg} Value Significance for Attribute/Metric Combinations Tested at a 5% Level of Significance*

PM	GGD	# of AFVs	VMT	# of Cust	# of Stations	# of Outreach	Stat. Density	Corridor Gap
Appropriateness	YES	YES	YES	YES	YES	YES	NO	YES
Clarity	YES	YES	YES	NO	YES	YES	NO	NO
Attainability	YES	YES	YES	YES	YES	YES	YES	YES
Objectivity	YES	YES	YES	YES	YES	YES	NO	NO

that allows for the appropriate application of the chi-square test to r_{wg} values. These acceptance values were used in this research for acceptance at the 5% level of significance.

5. DISCUSSION OF METRIC UNCERTAINTY ATTRIBUTES

The results shown in Tables 6-8 display the attribute/metric relationships in terms of raw agreement among survey respondents, the level of inter-rater agreement on that rating, and the statistical significance associated with each relationship. In the next sections each of these relationships we discuss these factors of metric uncertainty in relation to each of the 8 metrics of interest.

5.1 Gallons of fuel displaced (GFD)

The metric "Gallons of fuel displaced" or GFD is one of the most common metrics used to gauge the success and effectiveness of AF deployment grant programs. In the survey, 42 ratings were accumulated to assess uncertainty related to this metric; the highest number of ratings provided for any of the metrics tested. Tables 7 and 8 show that all metric/attribute relationships were found to be statistically significant. However, the strongest agreement among respondents for these relationships was for the appropriateness and clarity attributes. The metric GFD was found to be one of the most appropriate metric for discerning project-level differences between competing AF deployment projects (93% agreement and an r_{wg} score of 0.75). In addition, this metric showed similar scores for the attribute of clarity, indicating that the definition of the unit and the boundaries conditions surrounding that unit are well defined.

The attributes of data attainability and the level of objectivity associated with reporting when using this metric were found to have lower, but acceptable levels of agreement (r_{wg} of 0.62 and 0.64 respectively). This indicates that there may be some room for improvement when considering accessibility to the data needed to report the performance of projects using this metric and improving the level of objectivity that must be used when reporting performance using this metric. With regard to the attribute of data attainability specifically, the lower percentage of respondents giving a high rating in this category could be related to the confidentiality problems associated with accumulating throughput data from private refueling stations. It is common for stations to view giving this information as a potential loss to competitive advantage. Therefore, there is a need for developing applications and policies that allow these data to be gathered anonymously and confidentially so that data related to this metric become more readily available.

5.2 Number of AFVs purchased/subsidized

Similar to the results for the metric GFD, a high response rate was accumulated for this metric (n=36) and the results related to this metric were positive. While agreement levels were not as high for the appropriateness category, the capability of this metric for discerning and selecting among competing AF deployment projects was generally accepted among the survey population. The clarity of the definition of the unit and the boundary conditions surrounding the unit was also found to be high, indicating that the metric is simple and well understood. Additionally, survey respondents reported that data needed to report project-level performance using this metric were attainable and that little objectivity (human judgment) was required when calculating a value associated with this metric.

5.3 Vehicle-miles-traveled (VMT) with AFVs purchased

The vehicle-miles-traveled (VMT) metric is a common metric used in transportation research. Most commonly VMT is used as a gauge to measure growth or reduction of travel patterns for a population or geographic region (Jeon, et al., 2013). In this case, VMT relates to the use of AFVs once purchased (and not VMT reduction). Results for the VMT metric were found to be less positive and less conclusive than the results for the metrics previously discussed. While a relatively high response rate (n=32) was accumulated for this metric, tested attribute/metric relationships associated with the VMT metric were not found to be as high as for the metrics previously discussed. Within-group agreement values (r_{wg}) were low, although the hypothesis testing still shows significance beyond chance agreement among the ratings provided.

The appropriateness of this metric for discerning project-level differences among AF deployment projects showed lower than acceptable values for within-group agreement (r_{wg}=0.50). In addition, only 69% of respondents rated this metric as highly appropriate for discerning among competing AF deployment projects. Approximately 22% of respondents gave a neutral rating and 9% gave rated this metric as inappropriate for discerning project-level differences among competing AF deployment projects.

Respondents indicated that the clarity of the metric was relatively high although the agreement levels were acceptable but not strong (r_{wg} = 0.65). We suspect this result relates to the boundary conditions of the metric and not to the definition of the metric itself. The attainability and objectivity of this metric were also found to be quite low in terms of both the agreement in ratings among respondents and the related r_{wg} values (0.54 and 0.46 respectively).

While the definition of the unit is relatively well-understood (e.g. the clarity of the unit), there was a great deal of variance among the response relating to the attainability and objectivity of data reported for this metric. Approximately 45% of the respondents rated this metric as neutral or low in terms of data attainability. Similarly, 40% of respondents rated this metric as neutral or low in terms of the level of human judgment that must be used to report on values for this metric. Since the within-group agreement (r_{wg}) values are below acceptance thresholds, we cannot say with authority that the attainability or objectivity of the metric is neutral or low. However, it can be stated that disagreement does exist for these attribute/metric relationships beyond what could be expected from chance alone. Therefore, these results should provide the impetus for further research in these areas.

5.4 Number of customers served by AF refueling stations

When considering this metric a relatively low number of responses was accumulated in comparison to the response for other metrics (n=18). Similar to the VMT metric, several of the attribute categories displayed less than acceptable levels of within-group agreement (r_{wg} values) with the exception of the objectivity attribute. In the appropriateness category two-thirds of respondents rated this metric as highly appropriate while one-third of respondents gave it a neutral or low rating in this attribute category. Again, the interpretation of these results is not that the appropriateness of this *is* neutral or low. However, this does mean that there is disagreement on the appropriateness of this metric beyond what can be expected by chance alone, with a significant proportion of respondents indicating a rating other than high. The same is not true when considering the clarity attribute. In this case the hypothesis testing associated with the r_{wg} value shows that the responses are not beyond what can be expected from chance alone. Therefore there can be no conclusions drawn on this attribute/metric relationship.

Similar to the appropriateness attribute, data attainability shows a low amount of agreement, but agreement beyond what could be expected from chance alone. Notably, 72% of respondents rated the data attainability of this metric as neutral or low. Moreover, 56% rated data attainability as low performing. This is a strong indication that data collection and research efforts should be focused on how to collect these data. Similar to the GFD metric, there are sometimes problems with getting these data due to potential breaches of privacy and the potential for loss of competitive advantage that is sometimes perceived by refueling station owners.

5.5 Number of AF refueling stations constructed

The number of AF refueling stations installed by a project is a seemingly straight-forward metric. In most attributes/metric relationships some of the highest within-group agreement ratings were noted. The appropriateness category showed the lowest amount of agreement with 74% of survey respondents giving this metric a high rating. This may be a function of the use of this metric within wide-scope (i.e. fuel- and project neutral) project selection. When selecting among fuel-neutral project options using this metric, the selection process will almost inevitably select a plug-in electric project as the cost of installing recharging stations is only a fraction of the cost of installing refueling stations for other fuel types (e.g. such as a compressed natural gas (CNG) station for example). In terms of the attributes of clarity, data attainability, and objectivity high ratings and high levels of agreement were realized in all cases.   Like GFD, this indicates that the use of this metric is for measuring project-level effectiveness is recommended.

5.6 Number of outreach/education events

While the goal and subsequent measurement of contributing to AF deployment goals through outreach and education events was prevalent among survey respondents (n=37), the agreement among the attributes of this metric related to its' uncertainty were some of the lowest of all attribute/metric relationships tested. In general, we cannot say with any certainty that these attributes are low performing per se. However, the results of the hypothesis testing show that the variance in the responses is beyond what could be expected by chance alone. Moreover, the high levels of agreement that resulted for other metrics (e.g. GFD, # of AFVs, # of stations, etc.) when presented with the same question but a different metric indicates that there is dissention among agencies that personnel that use this metric within the attributes presented in this study.

When considering the appropriateness of this metric for discerning project level differences among competing AF deployment projects, approximately half of the respondents rated this metric as neutral or low performing. This is a significant finding especially when considering that many respondents cited outreach and education efforts as one of the most important programmatic goals (discussed in detail in later sections). When considering the clarity and the data attainability related to this metric, approximately three-fourths of the respondents rated this metric as high performing while one-quarter were neutral or gave a low performance rating. When considering the level of objectivity that must be used when reporting values for this metric, 64% of respondents rated this metric as high performing (low levels of human judgment required) while 36% rated this metric as neutral or low performing. These findings indicate that the use of this metric in the project selection process or in reporting the performance of an AF deployment project may be ill advised. In addition, there is a stark need for better defining the metric or finding a more appropriate metric for measuring contributions towards outreach and education efforts.

5.7 AF refueling station density and reductions in corridor gap refueling distance

The final two metrics tested in this research are related to the previous metrics (e.g. # of refueling stations installed), but different in that they compensate for the geographic need and demand for refueling stations. Interestingly, few respondents reported using these metrics, yet many respondents commented that the metrics could potentially be quite useful for designing programs and selecting AF deployment projects in the future. The ratings given for these metrics by the few respondents who did report using them gave them high marks in the category of data attainability and consistently so. Additionally, high marks were also given to the appropriateness for reducing corridor gap refueling distance. Likely due to the extremely small sample size of responses for the clarity and objectivity attributes (n=5 and n=8 respectively) the high level of variance within the rest of the ratings givens for these metrics ultimately failed to reject the null hypothesis, indicating that the findings were not beyond what could be expected by chance agreement alone. Therefore these metric show potential, but more information ratings on their will be required to draw any comparisons of these metrics to other metrics in use.

6. RESPONSES FROM EXPERT PANEL INTERVIEWS

To help answer "how" and "why" these results occur and to help provide convergent validity to the findings, follow-up interviews with selected respondents were arranged. To provide the most insight in these areas experts within the pool of respondents were targeted for participation. While the term "expert" is subjective, past studies that use focus groups or expert panels have helped to refine how experts are identified and selected (see Hallowell and Gambatese, 2010;

Hallowell, 2008 for a detailed discussion on this topic). In this study expert panelist candidates were identified by having met at least 3 of the following criteria:

- 10 or more years of experience with AF deployment grant programs;
- Membership in a nationally recognized committee related to AF deployment;
- Writer or editor of a book, book chapter, or manual on the topic of AF deployment;
- Current or past position of leadership within an agency or organization tasked with deploying AF technologies (typically denoted by a job title of senior consultant, manager, director, etc.); and
- Experience in teaching or directing workshops on AF deployment.

From the participating respondents, there were 19 survey participants that fit the criteria of 10 or more years of experience with AF deployment programs. Seven (7) responded to the request for an additional interview and were found to meet the aforementioned selection criteria. A semi-structured interview protocol was developed that included both closed and open-ended questions designed to provide participants with latitude to respond to the questions posed.

The first set of questions pertained to the broadest findings of the research. Gallons of fuel displaced (GFD), number of AFVs distributed (purchased or subsidized), and number of stations were found to be the only metrics for which there was agreement among the respondents about the performance level of the metrics. In addition, the consensus was that these metrics were relatively high performing in each of the attribute categories tested. Among the expert panel participants there was consensus that the definition of the unit was adequately defined and there were sufficient boundary conditions (the clarity attribute) as the score showed.

When considering the attribute of appropriateness for use in selecting competing AF deployment projects, one panelist stated that # of AFVs and # of station metrics are fuel dependent which might explain why there was less agreement on the appropriateness of these 2 metrics in particular. Another discussion in portion of the expert panel interviews was the attainability of the data. For this attribute, # of AFVs and # of stations displayed both a high level of performance and a high level of agreement among the respondents. However, there was noticeably less agreement for the GFD metric. To explain this, 3 panelists stated that this finding most likely related to GFD as a measure of throughput for refueling stations. One panelist stated "…the reporting of throughput values at refueling stations need refinement so that private companies can report the number of gallons of fuel equivalent distributed but not lose competitive advantage in the market". A second panelist stated that "…refueling stations are not in the business of collecting data…they are in the business of selling fuel in a very competitive market". Therefore there is a need according to both the survey findings and the panelists for improving how GFD can be reported, especially in the context of refueling stations specifically.

Similarly, the survey showed that the objectivity (level of human judgment) used in reporting GFD was lower than the level of agreement for # of AFVs and # of stations. One panelist suggested that there often was too much human judgment required to report GFD, which relates back to the problem of data attainability. One panelist stated that "…in the private sector the terms of gallons of fuel sold is considered proprietary information…" again reinforcing that there is a need for better tools to anonymously and confidentially report throughput without losing competitive advantage.

The second set of questions for the panel related to the relatively low levels of agreement associated with the VMT metric. The clarity of the VMT metric was the only attribute category to show a satisfactory level of agreement among survey respondents. When presented with this finding, the expert panel disagreed that there was a problem with the clarity of the definition of the unit as this is one of the most common units used the transportation field. However, several panelists agreed that there is commonly a problem with the boundary conditions associated with the unit. For example, VMT related to alternative fuel deployment specifically commonly relates to VMT in a non-attainment zone so that environmental conditions within that zone are specifically addressed. However, VMT is most commonly a self-reported metric, therefore determining if VMT by AFVs were actually traveled in a non-attainment zone (e.g. the boundary conditions of the metric) can be a problem. One panelist stated "…some people are conscientious and some people are not as far as reporting VMT to agencies". Conversely, another panelist stated that they had not experienced that problem so this may only occur in isolated cases.

Related to the discussion on the clarity of the VMT metric were the very low levels of agreement found for the attribute categories of attainability and objectivity. In reference to both categories several panelists stated that there are problems with getting data and therefore higher levels of human judgment that must be used when reporting the related data, but only for non-fleet applications. To illustrate the differences one panelist stated "…VMT in fleets is a really good metric with very little human judgment involved, especially for heavy-duty vehicles….idle time and miles traveled can be acquired by dumping (sic) data off of the engine". In contrast another panelist stated "…getting beyond the fleet perspective and to the public use [using the VMT metric] is important…". Interestingly the panelist also stated "If we ever move to a per mile tax instead of a per gallon tax (on fuel) there will be monitoring systems that will come about to address this problem". Therefore it appears that building policy and technology to measure VMT in private vehicles is an important need and for several reasons.

Similarly the next set of questions posed related to the metric of number of customers served by AF refueling stations installed. The clarity of the metric was found to be so varied that there were no usable findings beyond what could be expected by chance. Appropriateness and data attainability attribute categories were found to have very low levels of agreement as well. In response to these findings one panelist stated that there is great disparity between monitoring systems that provide these data for fleets versus the general public. To this end one panelist stated "…we have an enormous way to go with the public…we have more vehicles now and are beginning to get infrastructure, but there is still a great deal of resistance from the public whereas there is not from fleets findings when it comes to (recording) the number of customers served by refueling stations".

One of the most interesting findings in the survey was the complete lack of agreement related to the metric of number of outreach events. Expert panelists (and survey respondents) both routinely stated that outreach and education about alternative fuels is one of the most fundamental needs in deployment efforts. Yet the impact is difficult to measure. The clarity of the metric appears to relate to both the definition of the unit and the boundary conditions in this case. One panelist stated "…there are no real definitions on what constitutes an event". The panelist went further in stating that "there is a need to standardize the (measurement of) activities in this area". Related to objectivity of the metric, one panelist stated "The objectivity in which it's (sic) reported is definitely a problem, but I do think it is an effective metric". Another panelist stated "We can count how many people come to our workshops or websites, but we have no way to count the impact". Panelists were generally in consensus in stating that there was a need for better evaluation capacity in this specific area in general.

The final set of questions posed to the panelists related to the metrics of station density and corridor refueling gap distance. For these metrics there was an exceptionally low response rate which indicates that these metrics are not commonly used and makes interpreting any findings for these metrics difficult. When presented with the findings related to these metrics, expert panelists were generally in consensus that these metrics have great potential and are very important. One panelist stated that "…every project should have an evaluation of these metrics if they are available". Another panelist stated that "…these metrics are highly appropriate and easily calculated…". Unfortunately there still appear to be some privacy issues related to data attainability of these metrics. One panelist stated "…we can get information on where these vehicles are when a rebate program is used…but there may be privacy issues when trying to find out where other (alternatively fueled) vehicles are". Therefore, these metrics show great promise but are currently underused. In addition, there may be privacy issues related to these metrics which may need to be resolved in order to implement their use.

7. CONCLUSIONS

This paper presents a methodology for describing qualitative factors related to performance metrics commonly used to measure the performance of AF deployment programs. More specifically this paper looks at metrics related to the performance of grant programs as they have a particularly observable set of circumstances in comparison to other incentive types. From the broadest perspective, the findings show that there are distinct differences in the needed attributes of performance metrics currently used in AF deployment programs. Specifically, we find that gallons of fuel displaced (whether gasoline or diesel), number of AFVs purchased or subsidized, and number of AF refueling stations are the most reliable metrics. Conversely, metrics such as vehicle-miles traveled (VMT), number of customers served (or using) AF refueling stations, and number of outreach events *appear* to be quite specific given that they seem to be highly quantifiable. However, there are underlying problems in how these metric are defined, how attainable the data are for these metrics, how objectively data can be reported when using these metrics, and if they are even appropriate for discerning differences at the project level.

In particular, the metric gallons of fuel displaced was found to be one of the most commonly used metrics for measuring project and program performance and one of the best metrics by the attributes tested in this study. While this metric is well proven there are still issues of data attainability and objectivity in reporting using this metric, especially when describing the use of vehicle and refueling stations in the private sector. Attaining throughput data from private stations described in GFD or gallons of fuel equivalent (GGE) remains a difficult task as there are concerns of privacy, loss of competitive advantage, and the cost of resources needed to record and report these data. Therefore there is a need for providing private refueling stations with the ability to report these data without suffering any market penalty, disadvantage, or cost.

Similarly, measuring vehicle-miles traveled (VMT) and the number of unique refueling station users and refueling patterns among private citizens remains complicated. Issues of privacy commonly do not allow for information to be gathered from this sector. The development of technology that will allow for data to be gathered and used anonymously will be helpful to practitioners in the industry by helping them better locate refueling stations in the future based on the travel and refueling patterns of private vehicle owners. Furthermore, incorporating policies that will allow for the implementation of such technology will also be needed.

The research also shows that there is a distinct need for better measuring the impact of outreach and education efforts with the public. The development of a means for gauging the relative impact of different outreach and education tools (e.g. webinars versus “ride-and-drive” events), especially with correlation to cost, could provide a strong contribution to the many agencies and programs that work to deploy alternative fuels. This topic is discussed more in depth in the future research section.

Finally, including geographic need for alternative fuel refueling stations, whether in urban or in highway corridor settings, in the project selection and performance reporting processes could vastly improve efforts to deploy alternative fuels. The results of the survey showed that these metrics are underused but have high potential when considering the attributes used to compare metric uncertainty in this research.

8. LIMITATIONS AND FUTURE RESEARCH

The fundamental premise of this research is that metrics used to gauge the relative performance of AF deployment projects should be appropriate, clear, objective, and feasible (i.e. data attainability). To test this, metrics commonly used in practice are tested against these attributes with two primary results considered: 1) the level of performance (e.g. low versus high) for each metric/attribute relationship; and 2) the degree of consensus (i.e. agreement) for that rating. The results of this research show that there are disparate levels of performance and consensus among different metric/attribute combinations. However, some of the metric/attribute relationships show that, while a majority of survey respondents may rate a metric as high performing, the level of consensus remains unacceptably low. In these cases the interpretation of results is that there is room for improvement in these metrics. Perhaps a stronger finding would be lower performance ratings with a high level of consensus. However, no such findings were discovered in this research.

A second limitation of this research is the definition used in the survey for the clarity attribute category. The definition used in this research was related to the clarity of the definition of the unit *and* the definition of the boundary conditions associated with the unit. In hindsight, and as was discussed in this paper, this category could have been broken down into 2 unique categories: clarity of the unit and clarity of the boundary conditions. In future studies, we recommend doing so as this will provide a higher degree of insight related to the strengths and weaknesses of each metric. For example, the clarity of the VMT unit was found to have a satisfactory level of agreement for drawing conclusions on the findings in this research. However, VMT is one of the most common metrics used in transportation. As the expert panel showed the clarity related to the definition of the unit is likely not the problem. Alternatively, the problem with defining the boundary conditions such as defining if the vehicle-miles traveled are in a certain non-attainment zone as opposed to outside of an attainment zone is the problem related to the clarity of the unit (i.e. the boundary conditions of the unit). Therefore in future research we suggest looking at these 2 topics in isolation. In addition, questions related to VMT could be divided between VMT measured for fleet vehicles and VMT measured for non-fleet (e.g. personal or residential use). This would help delineate some of the problems associated with the metric in terms of data attainability and feasibility.

CHAPTER REFERENCES

Alternative Fuels Data Center (AFDC), 2013. All Laws and Incentives Sorted by Type. United States Department of Energy, Alternative Fuels Data Center (online), retrieved April 13, 2013 from http://www.afdc.energy.gov/laws/matrix/incentive.

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

Bureau of Transportation Statistics (BTS), 2011. National Transportation Statistics 2011, United States Department of Transportation, BTS (online), retrieved July 7, 2011 from http://www.bts.gov/publications/national_transportation_statistics/pdf/entire.pdf.

Byrne, M.R., Polonsky, M.J., 2001. Impediments to Consumer Adoption of Sustainable Transportation – Alternative Fuel Vehicles. International Journal of Operations and Production Management, 21(12), 1521-1538.

Calciolari, S., Cantu, E., Fattore, G., 2011. Performance Management and Goal Ambiguity: Managerial Implications in a Single Payer System. Health Care Management Review, 36(2), 164-174.

Chun, Y.H., Rainey, H.G., 2005. Goal Ambiguity and Organizational Performance in U.S. Federal Agencies. Journal of Public Administration Research Theory, 15(1), 1-30.

Deutch, J.M., 2011. An Energy Technology Corporation Will Improve the Federal Government's Efforts to Accelerate Energy Innovation, Brookings Institution, Discussion Paper 2011-05, May 2011.

Dunlap, W.P., Burke, M.J., Smith-Crowe, K., 2003. Accurate Tests of Statistical Significance for rwg and Average Deviation and Interrater Agreement Indexes. Journal of Applied Psychology, 88(2), 356-362.

Finn, F.H., 1970. A Note on Estimating the Reliability of Categorical Data. Education and Psychological Measurement, 30, 71-76.

Gallagher, K.S., Collantes, G., Holdren, J.P., Lee, H., Frosch, R., 2007. Policy Options for Reducing Oil Consumption and Greenhouse-Gas Emissions from the U.S. Transportation Sector, Energy Technology Innovation Policy, Discussion Paper, July 2007.

Gupta, J., Olsthoorn, X., Rotenberg, E., 2003. The Role of Scientific Uncertainty in Compliance with the Kyoto Protocol to the Climate Change Convention. Environmental Science & Policy 6(6), 475–486.

Gwet, K.L., 2010. Handbook of Inter-Rater Reliability, 2nd Ed., Advanced Analytics, LLC, Gaithersburg, Maryland.

Hallowell, M., 2008. A Formal Model of Construction Safety and Health Risk Management. Ph.D. dissertation, Oregon State Univ., Corvallis, Oregon.

Hallowell, M.R., Gambatese, J.A., 2010. Qualitative Research: Application of the Delphi Method to CEM Research. Journal of Construction Engineering and Management, 136(1), 99-107.

Heimann, L., 1995. Different Paths to Success: A Theory of Organizational Decision Making and Administrative Reliability. Journal of Public Administration Research and Theory, 5(1), 45-72.

Howell, D. and Chelius, M., 1997. Alternative Fuel Vehicle Programs – Applicability of Government Incentives. Transportation Research Record, Volume 1587, 1-9.

Hurst, E.G., 1980. Attributes of Performance Measures. Public Productivity Review, 4(1), 43-49.

International Energy Agency (IEA), 2002. Deployment Strategies for Hybrid, Electric, and Alternative Fuel Vehicles, Annex VIII Final Report.
International Energy Agency (IEA), 2011. Technology Roadmap – Electric and Plug-In Hybrid Electric Vehicles, Updated June 2011.
James, L.R, Demaree, R.G., Wolf, G., 1984. Estimating Within-Group Interrater Reliability With and Without Response Bias. Journal of Applied Psychology, 69(1), 85-98.
Jeon, C.M., Amekudzi, A.A., Guensler, R.L., 2013. Sustainability Assessment at the Transportation Planning Level: Performance Measures and Indexes. Transport Policy, 25, 10-21.
Kaplan, R.S., Norton, D.P., 1992. The Balanced Scorecard – Measures that Drive Performance. Harvard Business Review, Jan/Feb 1992, 1-10.
Kennerley, M., Neely, A., 2002. A Framework of the Factors Affecting the Evolution of Performance Measurement Systems. International Journal of Operations and Production Management, 22(11), 1222-1245.
Lance, C.E., Butts, M.M., Michels, L.C., 2006. The Sources of Four Commonly Reported Cutoff Criteria: What did they really say? Organizational Research Methods, 9(2), 202-220.
LeBreton, J.M., Senter, J.L., 2008. Answers to 20 Questions About Interrater Reliability and Interrater Agreement. Organizational Research Methods, 11(4), 815-852.
Lee, H., 2009. Oil Security and the Transportation Sector, in: Gallagher, K.S. (Ed.), Acting in Time on Energy Policy. Brookings Institution Press, Washington, D.C.
Melaina, M., McQueen, S., Brinch, J., 2008. Refueling Infrastructure for Alternative Fuel Vehicles: Lessons Learned for Hydrogen – Workshop Proceedings, National Renewable Energy Laboratory (NREL), Proceedings NREL/BK-560-43669, July 2008.
Melendez, M., 2006. Transitioning to a Hydrogen Future: Learning from the Alternative Fuels Experience, National Renewable Energy Laboratory (NREL), Technical Report NREL/TP-540-39423, February 2006.
Mood, A., Graybill, F., Boes, D., 1974. Introduction to the Theory of Statistics, McGraw-Hill Book Company Inc., New York, New York.
National Academy of Sciences (NAS), 2005. Thinking Strategically: The Appropriate Use of Metrics for the Climate Change Science Program, Committee on Metrics for Global Change Research, Climate Research Committee, National Research Council, National Academies Press, Washington, D.C.
Nahorski, Z., Horabik, J., (2008). Greenhouse Gas Emission Permit Trading with Different Uncertainties in Emission Sources. Journal of Energy Engineering 134(2), 47–52.
National Renewable Energy Laboratory (NREL), 2011. Commercialization and Deployment at NREL – Advancing Renewable Energy and Energy Efficiency at Speed and Scale, NREL/MP-6A42-51947, May 2011.
Neely, A., Gregory, M., Platts, K., 1995. Performance Measurement System Design: A Literature Review and Research Agenda. International Journal of Operations & Production Management, 15(4), 80 – 116.
Pandey, S.K., Wright, B.E., 2006. Connecting the Dots in Public Management: Political Environment, Organizational Goal Ambiguity, and the Public Manager's Role Ambiguity. Journal of Public Administration Research Theory, 16(4), 511-532.
Parkinson, S., Begg, K., Bailey, P., Jackson, T., 2001. Accounting for Flexibility Against Uncertain Baselines: Lessons from Case Studies in the Eastern European Energy Sector. Climate Policy 1, 55–73.

Plevin, R.J., 2010. Life Cycle Regulation of Transportation Fuels. Ph.D. dissertation. University of California at Berkeley, Berkeley, California.
Sagar, A., Gallagher, K.S., 2004. Energy Technology Demonstration and Deployment, in: National Commission on Energy Policy (NCEP) Technical Appendix for Chapter 6: Developing Better Energy Technologies for the Future, in: Ending the Energy Stalemate – A Bipartisan Strategy to Meet America's Energy Challenges, December 2004.
Smith, P., 1990. The Use of Performance Indicators in the Public Sector. Journal of the Royal Statistical Society, 153(1), 53-72.
Sobin, N., Molenaar, K., Cahill, E., 2012. Mapping goal alignment of deployment programs for alternative fuel technologies: An analysis of wide-scope grant programs in the United States. Energy Policy, 51, 405-416.
Sperling, D., Gordon, D., 2009. Two Billion Cars: Driving Towards Sustainability, Oxford University Press, New York, New York.
Sperling, D., Yeh, S., 2010. Toward a Global Low Carbon Fuel Standard. Transport Policy, 17, 47-49.
Sperling, D., 1988. New Transportation Fuels – A Strategic Approach to Technological Change, University of California Press, Berkeley, California.
United States Energy Information Administration (USEIA), 2011. Alternatives to Traditional Transportation Fuels 2009, Released April 2011, USEIA (online), retrieved July 7, 2011 from http://www.eia.gov/renewable/alternative_transport_vehicles/pdf/afv-atf2009.pdf.
United States Government Accountability Office (USGAO), 2011. Opportunities to Reduce Potential Duplication in Government Programs, Save Tax Dollars, and Enhance Revenue. Report GAO 11-318SP, March 2011.
Wagner, S.M., Rau, C., Lindemann, E., 2010. Multiple Informant Methodology: A Critical Review and Recommendations. Sociological Methods and Research, 38(4), 582

CHAPTER 4: DEMONSTRATING EVALUATIVE RELIABILITY IN ALTERNATIVE FUEL DEPLOYMENT PROJECT SELECTION

ABSTRACT

The deployment of alternative-fuel (technologies) to replace traditional gasoline and diesel vehicle is a common goal of public and private organizations. While there is consensus on the need for AF technologies, there is much less consensus on how to execute a unified, consistent, and effective effort for deploying them. This study builds upon previous research that identified relative uncertainty among metrics commonly used to gauge progress towards technology transition and energy security goals. Specifically, metrics that exhibited low levels of metric uncertainty are tested including gallons of fuel displaced (GFD), number of alternatively-fueled vehicles (AFVs) purchased or subsidized, and the number of alternative fuel refueling stations constructed. Ideally, these metrics would be reliable (i.e. repeatable, reproducible, and unequivocal) in their measurement towards technology transition and energy security goals. To test this, ten (10) context-specific scenarios were developed and presented to an expert panel consisting of 6 experts with a combined 107 years of experience in the field of AF deployment. The results show that there is a high level of repeatability for the metric GFD. Lower levels of repeatability are present for metrics that look at the number of vehicles and refueling stations in isolation. The results also show that the reproducibility of these metrics is highly context dependent. Fuel types that are the most similar to their gasoline and diesel counterparts are more easily and consistently discerned as high-performing or not high performing. Conversely, identifying high performance using these metrics in advanced fuel types is less reproducible. Applying these metrics to gauge the relative effectiveness of advanced fuel types may therefore be inappropriate. High levels of equivocality exist among the units of number of vehicles and number of refueling stations constructed across all fuel types considered in this research. Lower levels of equivocality exist when the metric GFD is considered both in isolation and when

presented within a context that includes the number of stations and/or the number of vehicles. These results will help agencies measure the success of their AF programs and researchers better understand how progress towards the goals can be gauged.

1. INTRODUCTION

The deployment of alternative-fuel (AF) technologies for the on-road transportation sector has been and continues to be a goal sought by a diverse group of public and private agencies throughout the United States and the world. Despite continuous efforts towards the deployment of AFs over several decades, AFVs (that use AFs) still represent less than two percent of registered light-, medium-, and heavy-duty vehicles on the road today in the United States (United States Energy Information Administration (USEIA), 2013). This is, at least in part, a consequence of an inability to fund and gauge the relative potential of competing AF projects.

One of the common problems among these programs is a lack of consistency in what goals are being addressed when deploying AF technologies and how we measure progress towards those goals consistently. This common theme in the literature highlights the need for consistent messages, consistent goals, and consistent measures towards goals (United States Government Accountability Office (USGAO), 2011; Deutch, 2011; International Energy Agency (IEA), 2011; National Renewable Energy Laboratory(NREL), 2011; Sperling and Yeh, 2010; Sperling and Gordon, 2009; Lee, 2009; Melaina, et al., 2008; Melendez, 2006; Sagar and Gallagher,2004; International Energy Agency(IEA), 2002; Byrne and Polonsky, 2001; Howell and Chelius, 1997; Sperling, 1988). Yet there are few studies that provide any in depth methodological (i.e. empirical) study into these areas. One aspect of improvement in this area is to test the evaluative reliability of metrics that are commonly used to gauge progress within AF deployment programs. It is this specific area that is tested in this research.

The term evaluative reliability can refer to many items related to performance measurement, statistics, and quality science bodies of knowledge. In this research, evaluative reliability refers to the quantitative aspects of performance metrics used in gauging the relative performance of AF deployment projects. In this paper, the term refers specifically to the following three aspects of performance metrics:

- <u>Repeatability</u> – the degree to which the same AF deployment project can be identified as high performing multiple times over time by the same rater (i.e. test-retest).
- <u>Reproducibility</u> – the degree to which AF deployment projects previously identified as high performing by one rater are identified as high performing by different raters.
- <u>Equivocality</u> – the existence of a quantitative threshold that can be used to discern between high-performing projects and projects that are something "other than" high-performing.

While these particular metric attributes should not be considered as the only requirements of effective performance metrics, they constitute the most quantitative aspects from a comprehensive list of recommended performance metric attributes previously accumulated from the relevant literature in this area (Hurst, 1980; Smith, 1989; Kaplan and Norton, 1992; Neely, et al., 1995; Heimann, 1995; Kennerley and Neely, 2002; Behn, 2003).

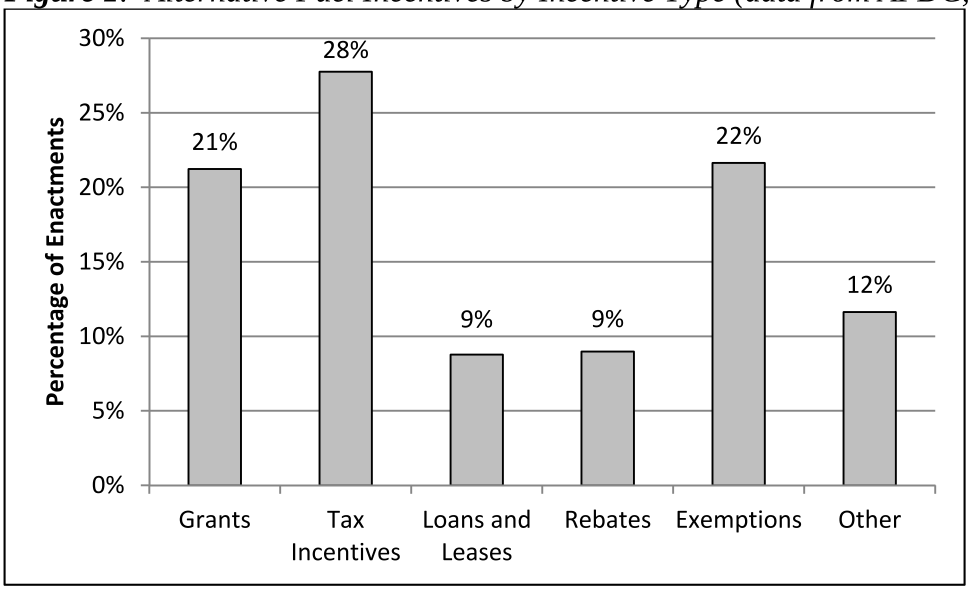
Previous research by the authors determined that three performance metrics show the greatest ability to for measuring progress towards technology transition and energy security goals; two common goals of AF deployment programs. These three metrics are: (1) gallons of fuel displaced (GFD) by a project; (2) number of AF refueling stations constructed; and (3) number of AFVs purchased or subsidized (see Chapter 3). The limited number of metrics tested is a function of the results from a previous study in this area. The authors tested several attributes of performance measures among commonly used metrics in AF deployment programs. Attributes tested included the clarity of the metric (related to the definition of the unit and the boundary conditions), the objectivity used in reporting the metric (i.e. the level of human judgment that is required), the degree of data attainability, and the appropriateness of the metric for discerning differences between competing projects. While common in use, metrics such as vehicle-miles traveled (VMT), number of outreach and education events, and number of unique customers using AF refueling stations exhibited low levels of clarity, objectivity, appropriateness, and data attainability. As a result, further study on the repeatability, reproducibility, and equivocality of these metrics (the attributes tested in this chapter) would be irrelevant as the associated numbers are incomparable. Therefore, the metrics tested in this study are the three that exhibited low levels of uncertainty.

2. GRANT PROGRAMS: AN OPPORTUNITY FOR AF DEPLOYMENT RESEARCH

One of the primary difficulties in studying AF deployment program evaluation systems is that few of them use a formal selection process to choose among competing AF types, vehicle platforms, and applications. Therefore it is difficult to discern any specific criteria on which competing projects are chosen. One type of incentive used to deploy AFs does use an observable system: grant programs. Grant programs typically use a systematic and observable approach to choose which projects to fund by documenting the performance towards specific metrics (e.g. tons of emissions reduced or GFD).

In addition to being an observable incentive type, grants are one of the more prevalent types of incentives used to deploy AFs. The Alternative Fuels Data Center (AFDC) keeps and updates a list of incentives for AF deployment (AFDC, 2013). Figure 1 shows the relative frequency with which each incentive type is used in 2013.

Figure 1*: Alternative Fuel Incentives by Incentive Type (data from AFDC, 2013).*



As can be seen in Figure 1, grants represent approximately 21 percent of incentives used to deploy AF technologies. Previous studies on grant programs show that there are many more grants than are included in the AFDC incentive database at state and regional levels (Sobin, et al., 2012).

In addition to being observable and prevalent, grant programs are also interesting as they are nearly a direct proxy for one of the more complex and vexing questions in this area: how do we select the most effective AF projects given the many choices among fuel type and vehicle platforms? Therefore, observing trends among projects funded by grants provides valuable insight towards the larger topic of how to evaluate AF deployment effectiveness in general.

3. RESEARCH APPROACH

To test evaluative reliability among performance metrics commonly used to gauge the performance of AF deployment programs we need to evaluate "real" numbers relating to the metrics. This research incorporates scenarios of real, high-performing projects that can be rated. In a previous study on AF project performance, a survey methodology was employed to collect data on the highest-performing, grant-funded projects in relation to several metrics commonly used to gauge progress towards energy security and technology transition goals (see Chapter 3). The response to the survey included 46 responses representing 29 percent of all Clean Cities Coalitions (CCCs) and 38 percent of state energy offices (SEOs) listed in the National Association of State Energy Officials (NASEO) database that utilize grant programs to deploy AF technologies. This provided a rich database from which to draw context-specific information from real, projects already dubbed the highest-performing, grant funded projects in relation to specific performance metrics.

The same study also gauged the relative metric uncertainty between metrics commonly used to gauge progress towards technology transition and energy security goals. In this study it was also discovered that only a few performance metrics related to technology transition and energy security goal domains displayed low levels of uncertainty such that a reliability study would be appropriate. As previously mentioned, metrics with appropriately low levels of metric uncertainty to be included in this study include GFD, number of AFVs purchased or subsidized, and number of AF refueling stations constructed.

To provide greater insight into the relative degree of evaluative reliability among these metrics, there must also be a context that relates to the many AF-types and vehicle platforms that are available within the on-road transportation sector. Including all of the possible matches of fuel type/vehicle platform combinations would be ideal. However, when considering data availability for high-performing projects related to the metrics of interest, it is not possible to test the evaluative reliability of all possible fuel-type and vehicle platform combinations. Therefore, a matrix was assembled to provide some degree of brevity in the number of possible combinations. The matrix combines similar AF-types based on the degree to which the physical characteristics of the AF vary from traditional gasoline or diesel platforms, the degree of additional difficulty for handling the fuel and installing a refueling infrastructure, and the relative cost difference for implementing the AF over traditional gasoline and diesel platforms. Based on these principles, a three-tiered matrix was constructed with each tier defined as follows:

- Tier I fuels – Hybrid technologies (including gas- or diesel-hybrid technologies), ethanol (E85), and biodiesel fuels;
- Tier II fuels – Liquefied petroleum gas (LPG), compressed natural gas (CNG), and liquefied natural gas (LNG); and
- Tier III fuels – Plug-in electric vehicles (PEVs) that use no type of combustion motor and hydrogen vehicles.

In general, tier I fuels are the lowest cost to implement and require the least amount of technological and behavioral change from the viewpoint of end users. In contrast, tier III fuels require higher costs to implement and require the greatest amount of technological and behavioral change from the end user. While combining fuel-types in the aforementioned tiers is somewhat subjective, it provides a basis on which scenarios can be tested.

Scenarios based on data from the previous survey on high-performing grant funded projects were incorporated into the tiered matrix. The resulting set of scenarios is shown in Table 1.

***Table 1**: Scenario Evaluation Matrix*

Fuel / Vehicle Platform Groups	Light- and Medium-Duty Vehicles	Heavy-Duty Vehicles
Ethanol (E85), Hybrid, Biodiesel (Tier I)	Scenario 1 - GGD per Vehicle (Light-Duty, Hybrid) Scenario 2 - GGD per Station (Light-Duty, E85)	Scenario 3 - GGD per Vehicle (Heavy-Duty, Hybrid)
LPG, CNG, LNG (Tier II)	Scenario 4 - GGD per Vehicle (Light-Duty, CNG) Scenario 5 - GGD per Station (Light-Duty CNG)	Scenario 6 - GGD per Vehicle (Heavy-Duty, LNG) Scenario 7 - GGD per Station (Heavy-Duty, CNG)
PEV, Hydrogen (Tier III)	Scenario 8 - GGD per Vehicle (Light-Duty, PEV) Scenario 9 - GGD per Station (Light-Duty, PEV, Highway Corridor) Scenario 10 - GGD per Station (Light-Duty, PEV, Urban Recharging)	*No Data*

The scenarios used in the study included project context and the quantitative data from each project related to the metrics tested (i.e. GFD, number of stations, and number of vehicles). Each scenario contained a brief description of the fuel-type, vehicle/station platform, vehicle/station application, the related GFD figure, and the related unit figure (i.e. number of stations or number of vehicles) for that project. For example, scenario 10 presented a project that would fund the installation of 250 recharging stations for light- and medium-duty PEVs in a densely populated downtown area, resulting in the displacement of 213 equivalent gallons of fuel per station per year and 53,300 gallons of fuel total per year (for the project). Each scenario incorporated data from the highest performing project for each matrix category that was attained from the survey.

Panelists were first asked to rate the scenario on a 5-point, interval scale on the overall contribution of the project towards technology transition and energy security goals. They were then asked to rate the individual aspects of the project with respect to the quantity of GFD, the number of vehicles purchased or subsidized by the project, and the number of AF refueling stations constructed using the same rating scale. Finally, each panelist was then asked to provide a numerical value for each quantitative scenario attribute (e.g. GFD/vehicle) that would have swayed their rating to the opposite end of the scale. This final set of questions was used to determine equivocality.

In total 10 scenarios were presented to the expert panel. For the light- and medium-duty PEV category two scenarios were included to gauge these attributes in relation to stations on highway corridors and stations installed for urban recharging. No examples of grant funded projects for heavy-duty, PEV or hydrogen applications were discovered in the survey. Therefore no scenario was included to represent this portion of the matrix.

The scenario-interview methodology utilized in this research incorporated separate interviews with experts in the field of AF deployment. While the qualification of "experts" in a field can be subjective, past studies and literature provide a roadmap for how objectivity can be maximized and bias minimized when using this type of methodology. Studies that use focus groups, nominal group techniques, or Delphi panels provide some guidance for or expert panels have helped to refine how experts are identified and selected (see Hallowell and Gambatese, 2010; Hallowell, 2008 for a detailed discussion on this topic). In this study expert panelist candidates were identified by meeting at least three of the following five criteria:

- 10 or more years of experience with AF deployment grant programs;
- Membership in a nationally recognized committee related to AF deployment;
- Writer or editor of a book, book chapter, or manual on the topic of AF deployment;

- Current or past position of leadership within an agency or organization tasked with deploying AF technologies (typically denoted by a job title of senior consultant, manager, director, etc.); and
- Experience in teaching or directing workshops on AF deployment.

From the participating respondents in the aforementioned study with 46 participants, 19 participants fit the criteria of 10 or more years of experience with AF deployment programs. Six (6) responded to the request for an additional interview and were found to meet at least three of the aforementioned selection criteria. In total the 6 respondents hold 107 years of experience in AF deployment programs, with an average of 18 years of experience per panelist.

4. RESULTS

The ratings provided by the expert are presented by the evaluative reliability attributes of repeatability, reproducibility, and equivocality. This section of the paper presents the results. The following section presents the discussion of the results.

Repeatability was tested by asking panelists to re-rate projects previously reported as the highest performing in the survey previously described. The expert panelists were not told that any of their own projects were included in the scenarios presented. On average, the time period between test/re-test was approximately two months. The two-month time period was adequate to determine if the participants could make their judgments repeatable. The results for this portion of the research are shown in Table 2 and are reported by metric in terms of raw agreement. A simple stoplight system that uses a red-yellow-green denotation is incorporated into the table. A green color indicates that the re-testing of metrics within the group were re-rated the same 100 percent of the time. A yellow color indicates that the repeatability of metrics was re-rated as high-performing at least 70 percent of the time. A red color indicates that the project was re-rated as high performing less than 70 percent of the time. While testing repeatability throughout the many contexts of the matrix would be ideal, the individual experience of panelists with every possible category within the matrix precludes this type of analysis. Therefore, the number of responses for each metric varies between four and six.

***Table 2**: Metric Repeatability*

Rating Category	Project Overall (n=5)	GFD (n=6)	GFD/Unit (n=6)	# of Vehicles (n=4)	# of Stations (n=5)
Low Performing	0%	0%	0%	25%	20%
Neutral	0%	0%	0%	0%	40%
High Performing	100%	100%	100%	75%	40%

Reproducibility was tested by asking panelists to rate projects previously reported as high-performing projects by other panelists. The raw agreement results are presented in Table 3 in terms of the relative frequency of ratings for each scenario (e.g. high, neutral, or low). Similar to the results presented in Table 2, the results presented in Table 3 utilize the same red-yellow-green stoplight system to denote areas of strength and areas of weakness when considering this metric. A green colored cell indicates that the project scenario was rated as a high-performing project at least 70 percent of the time and with no low ratings. A yellow colored cell indicates that the project received a high-rating less than 70 percent of the time, but was not given a low-performance rating 50 percent of the time or more. A red colored cell indicates that the project scenario was given a low-performing rating more than 50 percent of the time.

***Table 3**: Metric Reproducibility*

Fuel Platform Group	Scenario	Rating Category	Cumulative Scenario Metric Ratings: Project Overall	GFD/Unit	GFD	Unit (Station/Vehicle)
Ethanol (E85), Hybrid, Biodiesel (Tier I)	Scenario 1 - GFD per Vehicle (Light-Duty, Hybrid)	Low	0%	0%	0%	0%
		Neutral	0%	0%	0%	0%
		High	100%	100%	100%	100%
	Scenario 2 - GFD per Station (Light Duty, E85)	Low	17%	17%	0%	0%
		Neutral	17%	17%	0%	0%
		High	67%	67%	100%	100%
	Scenario 3 - GFD per Vehicle (Heavy-Duty, Hybrid)	Low	17%	17%	17%	0%
		Neutral	17%	17%	33%	33%
		High	67%	67%	50%	67%
LPG, CNG, LNG (Tier II)	Scenario 4 - GFD per Vehicle (Light-Duty, CNG)	Low	0%	17%	33%	33%
		Neutral	50%	17%	17%	67%
		High	50%	67%	50%	0%
	Scenario 5 - GFD per Station (Light Duty CNG)	Low	67%	67%	67%	33%
		Neutral	17%	17%	17%	50%
		High	17%	17%	17%	17%
	Scenario 6 - GFD per Vehicle (Heavy-Duty, LNG)	Low	0%	0%	0%	0%
		Neutral	0%	0%	0%	17%
		High	100%	100%	100%	83%
	Scenario 7 - GFD per Station (Heavy-Duty, CNG)	Low	17%	33%	33%	0%
		Neutral	17%	67%	50%	0%
		High	67%	0%	17%	100%
PEV, Hydrogen (Tier III)	Scenario 8 - GFD per Vehicle (Light-Duty, PEV)	Low	33%	33%	17%	17%
		Neutral	17%	17%	0%	0%
		High	50%	50%	83%	83%
	Scenario 9 - GFD per Station (Light Duty, PEV, Highway Corridor)	Low	50%	83%	83%	17%
		Neutral	17%	17%	17%	0%
		High	33%	0%	0%	83%
	Scenario 10 - GFD per Station (Light-Duty, PEV, Urban Recharging)	Low	50%	67%	50%	17%
		Neutral	0%	17%	33%	0%
		High	50%	17%	17%	83%

Table 4 presents the results for testing the equivocality of each metric both in isolation (i.e. units of refueling stations or number of vehicles individually) and in combination with the metric GFD. These results are presented in a dichotomous format (i.e. yes or no). A "yes" result indicates that all panelists agree that a quantitative threshold *does* exist that can discern between high-performing projects and projects that are something other than high performing. Conversely a "no" result indicates that at least one panelist indicated that no quantitative threshold could discern between high- and low-performing projects (i.e. the presence of equivocality – a negative attribute). While an actual threshold (i.e. a number) for each category would be ideal, these would likely be misleading and are not presented as there are several subcategories that could be included within each category of the matrix (e.g. heavy-duty refuse trucks versus heavy-duty long haul trucks).

***Table 4**: Metric Equivocality*

Fuel Platform Group	Scenario	Metric		
		GFD/Unit	GFD	Unit (Station/Vehicle)
Ethanol (E85), Hybrid, Biodiesel (Tier I)	Scenario 1 - GFD per Vehicle (Light-Duty, Hybrid)	YES	YES	YES
	Scenario 2 - GFD per Station (Light-Duty, E85)	YES	YES	YES
	Scenario 3 - GFD per Vehicle (Heavy-Duty, Hybrid)	YES	YES	NO
LPG, CNG, LNG (Tier II)	Scenario 4 - GFD per Vehicle (Light-Duty, CNG)	YES	YES	NO
	Scenario 5 - GFD per Station (Light-Duty CNG)	YES	YES	NO
	Scenario 6 - GFD per Vehicle (Heavy-Duty, LNG)	YES	YES	NO
	Scenario 7 - GFD per Station (Heavy-Duty, CNG)	YES	YES	NO
PEV, Hydrogen (Tier III)	Scenario 8 - GFD per Vehicle (Light-Duty, PEV)	YES	YES	YES
	Scenario 9 - GFD per Station (Light-Duty, PEV, Highway Corridor)	YES	YES	YES
	Scenario 10 - GFD per Station (Light-Duty, PEV, Urban Recharging)	YES	YES	NO

5. DISCUSSION OF EVALUATIVE RELIABILTIY AMONG METRICS

5.1 Repeatability of metrics

Table 2 shows that the expert panel was able to repeatedly rate their own projects (projects previously reported as the highest performing projects through the survey) as high performing, and the use of the vehicle or station as related to the GFD metric (GFD/unit) 100 percent of the time. The metric GFD in isolation was also re-rated as high performing 100 percent of the time, indicating that this metric is quite repeatable. The metric number of AFVs purchased or subsidized was re-rated highly 75 percent of the time indicating that repeatability is high with this metric as well, but with some disparity in the ratings. The low rating came in the vehicle-energy platform combination of light-duty, CNG vehicles (scenario 4 shown in the matrix). In general, the comment came from several raters that the "percentage of the fleet" would also influence their decision on how to rate the performance of projects using this metric. Table 2 also shows that the lowest amount of repeatability is associated with the number of AF refueling stations installed metric, with only 40 percent of the expert panel re-rating their own projects as high performing when considering this metric in isolation. More specifically, the lack of repeatability came in the area of CNG and LNG stations. While this finding may not be surprising, it is difficult to determine what a high versus a low number of CNG and LNG stations should be as there is often a low amount of station density to begin with and the capital cost for installing the stations is quite high.

5.2 Reproducibility of Metrics

When considering the overall project rating (the first column of ratings in Table 3), it is clear that there are significant differences between fuel-type categories. The first tier AF technologies such as ethanol, biodiesel, and hybrid technologies (the easiest and least expensive AF technologies to implement) are consistently rated the highest and are therefore the most reproducible. Conversely, second tier AFs such as LPG, CNG, and LNG (a greater degree of difficulty and cost to implement than first tier fuels) show lower levels of reproducibility indicating that the overall rating of these projects may be less reproducible and therefore less easy to identify. Low levels of reproducibility were discovered within the third tier AF technologies (the hardest degree of implementation difficulty). Overall, this indicates that identifying high-performing AF projects that utilize what may be deemed as advanced fuel types is more difficult than with other fuel types when considering the project as a whole. Similarly, these same trends are found when looking at the results for GFD per vehicle, GFD per station, and GFD as a metric in isolation.

When looking at the results of Table 3 for the unit in isolation (i.e. number of vehicles or number of stations), a different trend occurs. The expert panelists were more able to identify performance of higher tier AFs (i.e. PEVs) by the number of vehicles and the number of recharging stations. This finding is significant in that it shows us that that there is generally less knowledge surrounding what might constitute high-performance when considering the GFD metric as it is applied to more advanced fuel types.

This finding also highlights another very interesting topic: the ability of these metrics to describe and identify the potential performance of what could be very transformative projects. Scenario 9 (light-duty, PEV, highway-corridor) consists of a series of fast-recharging stations along a highway corridor. This scenario is almost a direct proxy for the Tesla Supercharger plan released in May of 2013 (see Tesla, 2013). While this project has potential to be one of the most transformative projects towards the goal domains of technology transition and ultimately energy security, the metrics most commonly used to measure the potential contribution of AF deployment projects are unable to detect or describe the potential of this project as is shown in the results of this portion of the tool. This highlights the need for improving the metrics in place and also for implementing newer metrics that are able to detect the potential of projects like this in the future.

5.3 Equivocality of Metrics

The final portion of results from this filter is testing the equivocality of these metrics. In this paper, we define equivocality as the acknowledgement that a numerical threshold exists that can delineate between high-performing projects and low-performing projects for each of the metrics, even if the numbers vary from one another. While it would be ideal to find a common (numerical) threshold for each context, the granularity of data related to the fuel-type/vehicle platform combinations used in the matrix is not fine enough to provide these thresholds. This topic is discussed at length in the limitations and future research section of this paper.

To execute this portion of the research, the expert panelists were asked to describe a threshold that would sway their rating towards the opposite end of the rating scale, regardless of their original response on the performance of the project towards energy security and technology transition goals. This technique allowed the panelists to indicate that a threshold for discerning between levels of performance does or does not exist. This form of inquiry in scenario analysis combines a few common expert panel inquiry methods: a forced choice response question with a "forced no-option" response, where refusal to answer is (in essence) the finding or result. That is, panelists give an initial rating for the performance of each project (scenario) presented. They are then asked to provide a number that would sway them towards the opposite end of the scale, without a "no response" option. If the panelist refuses to give a number (e.g. number of GFD or number of refueling stations) that would sway them towards a higher or lower performance rating this indicates that there is no numerical threshold (and therefore equivocality – a negative attribute).

Generally speaking, the results indicate that thresholds do exist (a positive attribute) for most categories. More specifically, the results indicate that there is low equivocality when considering the GFD metric and the use of this metric in relation to the units of stations or vehicles. Stated simply, this indicates that GFD both in isolation and in relation to other metrics and in different contexts is unequivocal (a positive attribute). That is, discernment between high and low performing projects is possible when using this metric.

Also notable are the areas where equivocality (a negative attribute) is distinctly present (denoted as “No” responses and highlighted in red). From a broad perspective, it is clear that problems of equivocality are present in only the units. That is, a discerning threshold does not exist for several of the contexts presented such as the number of vehicles or number of stations that would be representative of a high versus a low performing project. This phenomenon seems to occur primarily in the realm of heavy-duty vehicles platforms. As was previously discussed, the capital costs for purchasing vehicles in this area are very high. Discerning whether two versus three versus ten vehicles constitutes a high-performing project becomes very difficult. This problem is also true for discerning between high and low performing projects when considering the number of CNG and LNG refueling stations, as the capital costs are extremely high for these projects. The other specific area where equivocality appears to be a problem is in discerning what constitutes a high versus a low performing project in the area of number of recharging stations for PEV platforms, especially for urban applications (scenario 10). The reproducibility section of the larger evaluation capacity tool showed that the number of stations in isolation (instead of the GFD associated with vehicles and stations) were the most reproducible results when considering how performance in this specific part of the AF arena is measured. However, the equivocality results show that there may be problems associated with discerning high versus low performing projects using this metric in this context. Thus this finding highlights the need for greater evaluation capacity in this area specifically.

6. CONCLUSIONS

This paper presents a methodology for discerning evaluative reliability among different metrics that are commonly used to gauge project performance towards technology transition and energy security goals in AF deployment programs. We do so by testing specific attributes of evaluative reliability including repeatability, reproducibility, and equivocality. Literature in the area of performance measurement routinely states that metrics should be repeatable, reproducible, and unequivocal (Hurst, 1980; Smith, 1989; Kaplan and Norton, 1992; Neely, et al., 1995; Heimann, 1995; Kennerley and Neely, 2002; Behn, 2003). The results from this research show that there are significant differences within these attributes, especially when considering the many fuel types, vehicle platforms, and vehicle applications through which AF technologies are deployed.

The research shows that describing project performance using the GFD metric is highly repeatable when using the metric in conjunction with other metrics such as number of stations or vehicles (e.g. station throughput or vehicle consumption). In addition, the research shows that describing projects using the GFD metric is also highly repeatable when used to describe the relative performance of projects as a metric in isolation. Metrics such as number of AF refueling stations or number of AFVs show lower levels of repeatability. The metric number of AF refueling stations showed the lowest degree of repeatability.

When reproducibility is considered in the context of varying categories of fuel types and vehicle platforms, the research shows that lower-tier fuel types (fuel technologies that require the least amount of technological and behavioral change) such as hybrid technologies and ethanol/biodiesel fuel types are the most reproducible. Conversely, discerning high-performance among higher-tier fuel types (fuel technologies that require the greatest amount of technological and behavioral change) is more difficult. This finding pertains to both the metrics of interest as they are related to the project, as well as the overall potential of projects without consideration for the metrics. For practitioners and policymakers, this finding indicates that the performance metrics criteria included in this study may be less able to identify more advanced fuel types. Moreover, when given the chance to rate some of the most prolific projects in AF deployment for advanced fuel types without consideration of the metrics (the overall project rating) panelists still rated these projects as low performing 50 percent of the time. This finding highlights the need for metrics that can discern highly transformational projects for advanced fuel types.

The research also shows that the equivocality of metrics, when considered in the context of varying categories of fuel types and vehicle platforms does not show distinct trends related to these contexts. However, the equivocality is the highest (a negative attribute) when metric such as number of AF refueling stations and number of vehicles purchased or subsidized are presented in isolation. This indicates that using criteria such as the number of AF refueling stations constructed or the number of vehicles purchased as selection criteria for choosing among competing projects may be ill-advised as the meaning of high-performance is ambiguous without greater context.

7. LIMITATIONS AND FUTURE RESEARCH

While this research demonstrates the relative degree of evaluative reliability among different performance metrics and in different fuel-type/vehicle platform contexts, it does so through a methodology that elicits data from a small population (i.e. an expert panel). The use of this methodology is a function of the limited amount of data available to define high-performing projects needed to produce scenarios. While this methodology has provided insight into the topic of evaluative reliability and in which contexts (e.g. fuel-type/vehicle platform combinations) it is present, a "large-*n*" study might provide more definitive findings discovered in this research. Therefore we suggest such a study for future research, should the data become available to do so.

To determine evaluative reliability a set of scenarios was constructed using survey data from the highest performing projects from a large population of Clean Cities Coalitions and state energy offices. The highest performing project from this survey data was used to construct the scenario relevant to each unique section of the tiered fuel-type/vehicle platform matrix used in this research. While we are confident that each project represents a high-performing project, future studies could incorporate a low, medium, and high score to better determine the relative reproducibility and equivocality of the metrics tested.

Finally, this research presents results related to equivocality of metrics in a dichotomous context of existing or not existing. While the fuel-type/vehicle platform matrix incorporated in this study can discern trends in equivocality, a useful extension of these findings would be a quantitative range that defines what "is" and "is not" high performing when considering a finer degree of granularity. Future studies that could provide such a range for fuel-type/vehicle platforms such as what high-performance throughput for an LNG station for refuse trucks versus an LNG station for long-haul trucks could prove to be quite useful in the project selection process. As a limitation of this research we could not provide this degree of certainty due to the lack of available data.

CHAPTER REFERENCES

Alternative Fuels Data Center (AFDC), 2013. All Laws and Incentives Sorted by Type. United States Department of Energy, Alternative Fuels Data Center (online), retrieved June 10, 2013 from http://www.afdc.energy.gov/laws/matrix/incentive.

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

Bureau of Transportation Statistics (BTS), 2011. National Transportation Statistics 2011, United States Department of Transportation, BTS (online), retrieved July 7, 2011 from http://www.bts.gov/publications/national_transportation_statistics/pdf/entire.pdf.

Byrne, M.R., Polonsky, M.J., 2001. Impediments to Consumer Adoption of Sustainable Transportation – Alternative Fuel Vehicles. International Journal of Operations and Production Management, 21(12), 1521-1538.

Deutch, J.M., 2011. An Energy Technology Corporation Will Improve the Federal Government's Efforts to Accelerate Energy Innovation, Brookings Institution, Discussion Paper 2011-05, May 2011.

Gallagher, K.S., Collantes, G., Holdren, J.P., Lee, H., Frosch, R., 2007. Policy Options for Reducing Oil Consumption and Greenhouse-Gas Emissions from the U.S. Transportation Sector, Energy Technology Innovation Policy, Discussion Paper, July 2007.

Hallowell, M., 2008. A Formal Model of Construction Safety and Health Risk Management. Ph.D. dissertation, Oregon State Univ., Corvallis, Oregon.

Hallowell, M.R., Gambatese, J.A., 2010. Qualitative Research: Application of the Delphi Method to CEM Research. Journal of Construction Engineering and Management, 136(1), 99-107.

Heimann, L., 1995. Different Paths to Success: A Theory of Organizational Decision Making and Administrative Reliability. Journal of Public Administration Research and Theory, 5(1), 45-72.

Howell, D. and Chelius, M., 1997. Alternative Fuel Vehicle Programs – Applicability of Government Incentives. Transportation Research Record, Volume 1587, 1-9.

Hurst, E.G., 1980. Attributes of Performance Measures. Public Productivity Review, 4(1), 43-49.

International Energy Agency (IEA), 2002. Deployment Strategies for Hybrid, Electric, and Alternative Fuel Vehicles, Annex VIII Final Report.

International Energy Agency (IEA), 2011. Technology Roadmap – Electric and Plug-In Hybrid Electric Vehicles, Updated June 2011.

Kaplan, R.S., Norton, D.P., 1992. The Balanced Scorecard – Measures that Drive Performance. Harvard Business Review, Jan/Feb 1992, 1-10.

Kennerley, M., Neely, A., 2002. A Framework of the Factors Affecting the Evolution of Performance Measurement Systems. International Journal of Operations and Production Management, 22(11), 1222-1245.

Lee, H., 2009. Oil Security and the Transportation Sector, in: Gallagher, K.S. (Ed.), Acting in Time on Energy Policy. Brookings Institution Press, Washington, D.C.

Melaina, M., McQueen, S., Brinch, J., 2008. Refueling Infrastructure for Alternative Fuel Vehicles: Lessons Learned for Hydrogen – Workshop Proceedings, National Renewable Energy Laboratory (NREL), Proceedings NREL/BK-560-43669, July 2008.

Melendez, M., 2006. Transitioning to a Hydrogen Future: Learning from the Alternative Fuels Experience, National Renewable Energy Laboratory (NREL), Technical Report NREL/TP-540-39423, February 2006.

National Academy of Sciences (NAS), 2005. Thinking Strategically: The Appropriate Use of Metrics for the Climate Change Science Program, Committee on Metrics for Global Change Research, Climate Research Committee, National Research Council, National Academies Press, Washington, D.C.

National Renewable Energy Laboratory (NREL), 2011. Commercialization and Deployment at NREL – Advancing Renewable Energy and Energy Efficiency at Speed and Scale, NREL/MP-6A42-51947, May 2011.

Neely, A., Gregory, M., Platts, K., 1995. Performance Measurement System Design: A Literature Review and Research Agenda. International Journal of Operations & Production Management, 15(4), 80 – 116.

Sagar, A., Gallagher, K.S., 2004. Energy Technology Demonstration and Deployment, in: National Commission on Energy Policy (NCEP) Technical Appendix for Chapter 6: Developing Better Energy Technologies for the Future, in: Ending the Energy Stalemate – A Bipartisan Strategy to Meet America's Energy Challenges, December 2004.

Smith, P., 1990. The Use of Performance Indicators in the Public Sector. Journal of the Royal Statistical Society, 153(1), 53-72.

Sobin, N., Molenaar, K., Cahill, E., 2012. Mapping goal alignment of deployment programs for alternative fuel technologies: An analysis of wide-scope grant programs in the United States. Energy Policy, 51, 405-416.

Sperling, D., Gordon, D., 2009. Two Billion Cars: Driving Towards Sustainability, Oxford University Press, New York, New York.

Sperling, D., Yeh, S., 2010. Toward a Global Low Carbon Fuel Standard. Transport Policy, 17, 47-49.

Sperling, D., 1988. New Transportation Fuels – A Strategic Approach to Technological Change, University of California Press, Berkeley, California.

Tesla Motors, Inc., 2013. Supercharger: The fastest recharging station on the planet, Tesla Motors, Inc., (online), June 10, 2013 from http://www.teslamotors.com/supercharger.

United States Energy Information Administration (USEIA), 2013. How many alternative fuel and hybrid vehicles are there in the U.S.?, USEIA (online), June 10, 2013 from http://www.eia.gov/tools/faqs/faq.cfm?id=93&t=4.

United States Government Accountability Office (USGAO), 2011. Opportunities to Reduce Potential Duplication in Government Programs, Save Tax Dollars, and Enhance Revenue. Report GAO 11-318SP, March 2011.

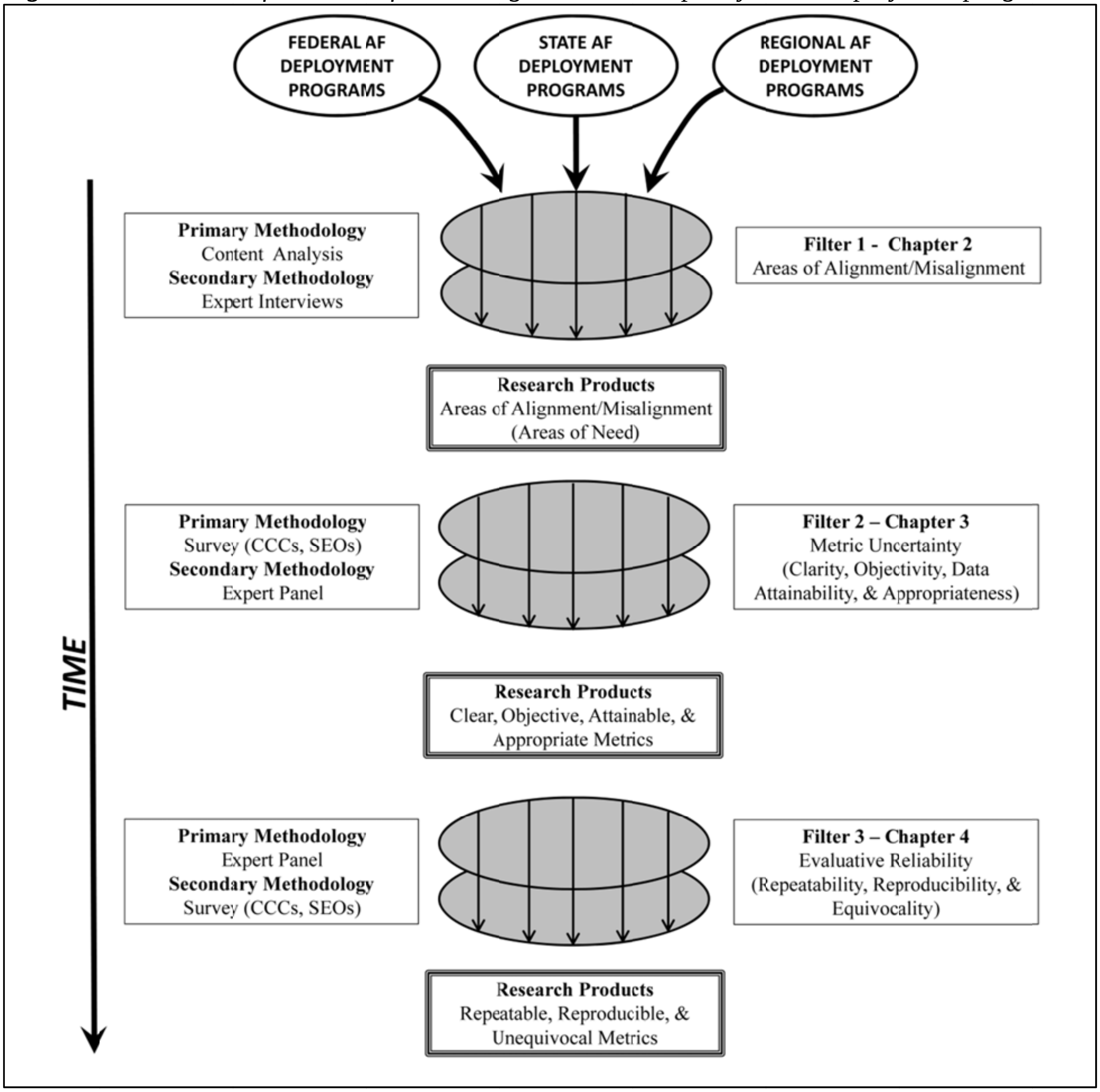
CHAPTER 5: CONCLUSION

CONTRIBUTIONS

The three papers included in this dissertation collectively contribute to the body of knowledge in evaluation capacity building for AF deployment programs. They conduct a thorough investigation in areas of AF deployment and provide new framework for a comprehensive view of how and where federal, state, and regional agencies can gain evaluation capacity to improve performance. The first paper provides a framework and state-of-practice analysis on what goals are being addressed and by whom when agencies use wide-scope grant programs to deploy AF technologies. The second paper comparatively analyzes qualitative factors of uncertainty among metric commonly used to gauge progress towards technology transition and energy security goals specifically. The third paper presents a comparison of the reliability for the three metrics that were deemed to be the most appropriate to gauge progress towards technology transition and energy security goals.

Collectively, these three papers and methodologies can be thought of as a set of filters (an evaluation framework) through which effective metrics must pass (metaphorically analogous to a sieve analysis). This framework is applicable to metrics currently in use, but could also be used to assess the potential of metrics in the future. Figure 1 presents the framework in its entirety. Figure 1 includes each of the methodologies utilized for each filter and the resulting research products.

***Figure 1**: The holistic framework for building evaluation capacity in AF deployment programs*



The systematic approach to building evaluation capacity in the area of AF deployment programs addresses two observed problems: one practical and one theoretical. First, there is a practical problem related to how and why "we" deploy AF technologies in the United States. The multiple and often redundant efforts to deploy AFs are performed at several different governmental levels. These deployment efforts exhibit varying degrees of alignment (and misalignment) in terms of both what goals are addressed and how performance towards those goals is evaluated. In the United States alternatively fueled vehicles (AFVs) still represent less than two percent of registered vehicles (USEIA, 2013) on the road despite decades of deployment efforts. This is, at least in part, a result of the misalignment and general lack of synergy and cooperation among agencies that seek to deploy AFs.

There is also a theoretical problem related to how to gauge the coherency and reliability of performance measurement systems, both now and in the future. While there is a wealth of literature that describes the *need* for common attributes of performance measurement systems under the broader heading of evaluation capacity building (Schaumburg and Muller, 1996; Bamberger, 2000; Stockdill, et al., 2002; Clewell and Campbell, 2008;), and more specifically the metrics used in these systems (Hurst, 1980; Smith, 1990; Kaplan and Norton, 1992; Neely, et al., 1995; Heimann, 1995; Kennerley and Neely, 2002; Behn, 2003), there are few if any methodologies that address *how* to systematically evaluate these attributes. As a result, performance metrics used to evaluate AF deployment programs, as well as many other sets of programmatic metrics that attempt to measure progress towards goals in similar settings, are commonly created and implemented in a somewhat *ad hoc* manner with very few checks and balances installed to see if they are providing usable, relevant information.

CONTRIBUTIONS TO PRACTICE

The research and findings contained in this dissertation contribute to practice in the AF deployment arena in several ways. Public and private agencies that work diligently to deploy

AFs are the primary benefactors of this research. One of the first points of departure of this research is the construction of a fundamental application for evaluating the alignment of programs. Chapter 2 of this dissertation was executed by applying the theories of goal and evaluation ambiguity as first described and implemented by Chun and Rainey (2005) to the AF deployment arena from a holistic viewpoint. I am aware of no other studies that have attempted to look at AF deployment efforts as a whole as I have in Chapter 2. Perhaps unsurprisingly, the findings indicate that there is a great deal of alignment towards environmental goals. I state that this finding is “unsurprising” since environmental goals are, at least anecdotally (before this research), one of the most common reasons we hear of the need for implementing AFs. Therefore the first practical contribution is the creation of a methodology for “goal mapping” as is described in Chapter 2. This part of the framework helps us define “why” we deploy AFs. In Chapter 2 the research also shows that the direct quantitative measurement of environmental benefit is the most common metric by which we “choose” among competing AF deployment projects. Again, this finding may be somewhat “unsurprising” as the practical contribution is an incremental advancement, but there is no previous research or literature that addresses this topic exclusively.

The larger practical contribution of Chapter 2 is the definition of a simple, practical means for evaluating these types of programs from a holistic perspective and gleaning information on the areas of relative alignment and misalignment. If this type of evaluation is implemented in an iterative (time-series) manner and compared to the relative outcomes of these programs over time, it would provide a valuable (and practical) steering mechanism for the iterative design of policy so that the relative effect of including and excluding certain elements in the project selection process can be ascertained. Based on the summative findings of this research, I surmise that including and selecting AF deployment projects based on alternative metrics (indirect measures) of technology transition *could* result in greater reductions of environmental pollutants and gallons of traditional fuels consumed (discussed in greater detail in the "future research" section).

Chapter 3 of this dissertation presents the implementation of a survey methodology that can be used to measure the relative ability of metrics to correctly and accurately communicate a level of progress towards technology transition and energy security goals under the broader heading of metric uncertainty. This methodology utilizes a simple rating system for evaluating predefined, qualitative attributes of the metrics used to select among competing projects and report the relative performance of them. The technology transition and energy security goal domains became the focus of the chapter as they were, collectively, the areas of greatest misalignment and therefore the areas of greatest research need.

The practical contributions of this chapter come primarily from the findings in these areas of misalignment. The findings indicate that, when compared side-by-side, there are distinct patterns and differences among performance metrics commonly used in these areas. Using our filter analogy (see Figure 1) this filter catches the largest problems with performance metrics: the clarity of the definition and boundary conditions of the metric, the attainability of the data, the objectivity (i.e. level of human judgment required) with which data are reported, and the appropriateness of the metric for discerning differences between competing projects. The findings show that GFD, number of refueling stations constructed, and number of AFVs purchased/subsidized are the only metrics that pass this filter. That is, when the qualitative factors of effective metrics are considered, these are the metrics that are high performing. Conversely, VMT, the number of customers served by AF refueling stations, and the number of outreach/education events were found to be low performing in several areas and therefore did not pass this filter. The practical implications of this finding are significant. There is a strong indication that the metrics not passing this filter are not useful for selecting among competing AF deployment projects. These findings can be directly applied by the many public and private agencies that work to deploy AFs when developing new programs and selection processes for those programs. Moreover, in Chapter 2 we showed that these underperforming metrics are currently used to select among competing projects and describe the relative performance of these projects. This indicates that when numbers are presented related to these metrics there is a high likelihood that they are invalid in that they are not objective and may or may not be measuring the same thing (e.g. incomparable data).

The other practical contribution of this portion of the research is the identification of areas of need. From a technological development perspective this research outlines two very specific areas of need:

1. the need for an *automated and anonymous* way for stations to report throughput and the number of unique users at AF refueling stations without losing competitive advantage; and
2. the need for an *automated and anonymous* way to collect and report VMT and travel use patterns from private vehicles.

From a methodological perspective Chapter 3 also shows a distinct need for creating non-intrusive ways for measuring the effect (i.e. impact) of outreach and education efforts geared towards the public. For example, providing agencies with a roadmap for where to most effectively spend their next outreach dollar would be an invaluable contribution in this area. The importance for further research in this area cannot be overstated as informing the public on the needs, benefits, and drawbacks of implementing AF technologies will have to occur before major technology transition can occur.

The last practical contribution of this portion of the research is the identification of potential for implementing selection criteria that include geographic need for the location of refueling stations. Some programs (and states/regions) have been highly successful at doing this while others (most) still do not include this as a consideration for funding a potential project. The expert panel that was incorporated in this study to provide convergent validity for this section of the research unanimously indicated that these metrics are important and should be incorporated in the project selection process. Moreover, correlating potential station location to a predictable, optimized throughput (i.e. GFD) for different areas would be an ideal application for these metrics in the project selection process.

Chapter 4 (the third paper included in this dissertation) provides several contributions to practice through further analysis of the metrics that passed the second filter in the framework, under the broader heading of evaluative reliability. In this part of the framework the repeatability, reproducibility, and equivocality of the metrics that passed the prior qualitative filter (see discussion on Chapter 3 findings) were tested through presentation of multiple scenarios and the subsequent rating of those scenarios by an expert panel.

Similar to Chapter 3, the practical contributions come primarily through the findings of the research. The primary practical contribution relates to the findings for the GFD metric; an extremely common metric used to gauge AF deployment performance. The research shows the while GFD has high levels of repeatability, the relative degree of reproducibility decreases with this metric as fuel types become more advanced. Stated simply, this demonstrates that metrics such as GFD may be inappropriate for gauging the relative potential of advanced fuel types, especially when compared on a cross-project basis to less advanced fuel types (e.g. ethanol or biodiesel). The potential ramifications of this finding are significant in that this provides some empirical evidence that using this metric in the project selection process is not necessarily part of a fuel-neutral approach.

The other practical contribution pertains to the reproducibility findings. The experts tend to think about project potential and performance for advanced fuel type in terms of number of stations and number of vehicles, and not GFD. This is in contrast to the same attribute in other fuel types where the GFD metric is well understood. When we combine this finding with the equivocality results that show that experts have the hardest time defining quantitative thresholds that define high performance in terms of units (i.e. number of stations or vehicles), this indicates that there really are no metrics that can consistently gauge the relative contribution of advanced fuel types towards technology transition and energy security goals.

The practical contribution of this is significant in that it informs practitioners and policymakers that gauging the relative performance of advanced fuel types with these metrics is ineffective. Other metrics are needed to gauge the relative contributions of projects that use advanced fuel types. To highlight this point, two of the scenarios presented to the expert panel participants were PEV recharging infrastructure projects. Of particular note is the scenario that included a series of 14 level 3 fast charging stations along a highway corridor. When given the opportunity to rate this project as a whole (without looking at the individual quantitative attributes of the project), half of the expert panel gave the project a low rating. When presented with the low throughput values (which likely will be accurate for the first several years after they are constructed) the expert panelists gave these numbers a low rating as well. This finding is particularly significant as this scenario could be construed as a direct proxy for the Tesla Supercharger Network (see Tesla, 2013). Said another way, the practical contribution of just this finding is that the metrics currently in place may not be able to detect what may be the most transformational projects in the industry. Therefore this finding illustrates the potential need for long-term project selection criteria and metrics that can compensate for this deficiency.

CONTRIBUTIONS TO THEORY

The research contained in this dissertation contributes to theory primarily by changing the way we think about evaluation and evaluation capacity both in the context of AF deployment programs and in the context of performance measurement systems in general. This research does so by contributing a combined framework for putting numbers to, what are in the literature, rather abstract needs in several areas. With respect to the AF deployment related body of literature, the need for consistent, aligned, and systematic approaches for defining and measuring progress towards goals is routinely stated (United States Government Accountability Office (USGAO), 2011; Deutch, 2011; International Energy Agency (IEA), 2011; National Renewable Energy Laboratory(NREL), 2011; Sperling and Yeh, 2010; Sperling and Gordon, 2009; Lee, 2009; Melaina, et al., 2008; Melendez, 2006; Sagar and Gallagher,2004; International Energy Agency(IEA), 2002; Byrne and Polonsky, 2001; Howell and Chelius, 1997; Sperling, 1988). However, there is a dearth of literature that broaches these topics empirically. With regard to the realm of performance measurement, the literature in this area routinely states that metrics should be qualitatively and quantitatively consistent in what metrics measure and the interpretation of data delineated by metrics (expressed as a list of several attributes) (Hurst, 1980; Smith, 1989; Kaplan and Norton, 1992; Neely, et al., 1995; Heimann, 1995; Kennerley and Neely, 2002; Behn, 2003; NAS, 2005). However, there has been no clear method on how to measure or quantify these attributes. This research creates a framework for measuring and quantifying these attributes so that the relative effectiveness of performance metrics and criteria can be compared and contrasted. The use of this system also provides a framework for finding where and how evaluation systems are sufficient (or deficient). This approach can be used to test both performance metrics that are in place, as well as, future performance metrics.

The primary theoretical contribution of Chapter 2 is the system-wide approach to analyzing AF deployment programs at many governmental levels and by extending theory on evaluative goal ambiguity. This research makes a theoretical contribution by looking at multiple AF deployment programs as a system rather than on a program-by-program basis as is currently and commonly done (see Johnson, 2012; Almeida, 2011; Johnson and Bergeron, 2008; FHWA, 2008A; FHWA, 2008B). In contrast I have assembled a high-level view of all programs that use AF deployment as a tool towards a myriad of goals. While this approach has its limitations, I am unaware of any other research program that has attempted to view the use of AF deployment programs as a whole as I have in Chapter 2. In addition, and by using this approach, this method allows for comparisons to be drawn between federal, state, and regional programs that use AF deployment as a tool. Theoretically, using this new perspective may allow future researchers in areas of economic and policy research to find areas of potential for greater synergy and ultimately an optimization level for that synergy that could lead to more effective deployment efforts of AF technologies.

The theoretical contribution from Chapter 3 (the 2nd paper presented in this dissertation) is the formation and execution of a methodology to evaluate the qualitative (if not abstract in some cases) attributes of performance metrics that literature routinely cites as necessary (Hurst, 1980; Smith, 1989; Kaplan and Norton, 1992; Neely, et al., 1995; Heimann, 1995; Kennerley and Neely, 2002; Behn, 2003; NAS, 2005). While the literature routinely states that there is a need for clarity, objectivity, data attainability, and feasibility among performance metrics, there is no literature or research that provides a systematic means for evaluating these attributes such as what I have provided. Moreover, there is no literature that does so in the area of AF deployment; one of the most important topics of our time.

The theoretical contribution from Chapter 4 (the 3rd paper presented in this dissertation) is the formation and execution of a methodology to assess the relative differences in evaluative reliability (e.g. repeatability, reproducibility, and equivocality) among different metrics of interest. A particular limitation of this research was a lack of available data. To execute this portion of the research a small population study was employed which yielded interesting and compelling findings related to the differences between metrics. For future research, a "large-*n*" study might be more definitive. However, until there is some reasonable level of deployment (especially for advanced fuel types), there will continue to be few data available. As a result, the methodology presented in this study could be continually used to assess evaluative reliability over time until data do become more readily available.

Holistically, the framework presented in this research under the broader heading of building evaluation capacity (see Figure 1), can and should be used to continually identify areas of evaluation need and to develop new metrics. Therefore there is a theoretical contribution in that the framework presented that can be used to help define new metrics to better identify the potential of transformative technologies and projects.

LIMITATIONS

While the findings from this research have provided a great deal of insight into how to measure the evaluation capacity of AF deployment programs, there are a number of limitations for this research that should be noted. First, cost was not incorporated in this study as costs were found to be considered proprietary information and was therefore inaccessible. Incorporating costs associated into the metrics studied in this research will improve the findings and usability of this research. While some AF deployment programs publish detailed expenditure data, most present this information in aggregate. Stated simply, disaggregating expenditures related to AF deployment requires a finer degree of granularity than is available to the general public. A more transparent approach for disseminating program and project cost performance may also have the unexpected effect of creating a more diverse and robust conversation on these topics. Pielke (2009) describes this as an "honest broker" approach and suggests that ultimately this approach can be more effective.

Secondly, I have attempted to view AF deployment programs holistically. That is, I have tried to include, as a single population, the vast number of programs that use AF deployment either as a tool or as part of a broader programmatic goal. While viewing AF deployment efforts in this way provides a high level view of where programmatic alignments and misalignments occur, there was no obvious way to weight the percentage that each program used AF deployment as a tool in comparison to other tools in the arsenal. This was again primarily due to a lack of data on programmatic costs and spending. Without cost figures it becomes difficult at best to describe the degree to which each program utilizes AF deployment towards their specific programmatic goals. Incorporating a programmatic weight by cost could provide greater insight and perhaps more accurately describe why and how “we” choose alternative fuels, as was the focus of this portion of the research.

In the portion of dissertation that describes metric uncertainty (Chapter 3) the attributes of clarity, objectivity, data attainability, and appropriateness were tested over several metrics. While a great deal of insight was gained from testing these attributes, future studies that utilize this approach may consider splitting a few of these categories so that greater insight may be achieved. For example, the clarity attribute, as it was defined in this paper, pertained to the definition of the unit measured and the boundary conditions of that unit. Dividing these into two separate topics could provide a greater degree of insight into the potential strengths and weaknesses of each metric. For example, the VMT metric showed acceptable, but not high, values of agreement in the clarity area. This finding was a bit surprising as this metric has been widely used and accepted in both transportation research fields and in industry. Through the expert panel incorporated in this study I found that the problems related to the clarity of this metric were a function of the lack of access to data on *where* the vehicle-miles were traveled. Therefore the definition of the unit was quite clear but the surrounding boundary conditions associated with the unit were not.

Considering the findings of the qualitative filter, there is also a distinct need for more research in the area of metrics that relate to the geographic need of stations, especially in relation to the optimization of throughput by strategically locating stations. The need for more information and the implementation of metrics that compensate for geographic need in the project selection process cannot be overstated.

Finally, the research related to evaluative reliability (Chapter 4) presented findings on the level of repeatability, reproducibility, and equivocality associated with specific metrics and in different contexts. For this portion of the research, we utilized an expert panel which provided a great deal of insight from people with a great deal of experience over time in AF deployment efforts. While the high levels of experience were powerful, the small number of participants may induce certain forms of cognitive bias. The impact of any bias can be exacerbated in small sample methodologies as the rating from one expert constitutes a large percentage of the results. Therefore we present the results of this section noting this caveat. In future studies, a larger sample size could provide a more comprehensive insight into these areas, especially in the area of repeatability.

In testing the reproducibility for identifying high-performing projects, I presented scenarios in the context of different fuel-type and vehicle-platform combinations. This approach helped identify the differences in reproducibility within the many fuel-type and vehicle-platform combinations that are available when deploying AFs. However, in each case we presented only one high-performing project for each context. Future studies in this area should consider incorporating several high-performing projects within each context to provide greater insight into the reproducible identification of high-performing projects. This approach will help determine what the collective threshold for high-performing projects may be. When testing the equivocality of each metric, we determined “if” and “where” equivocality exists, based on both metric and context. However, using several scenarios within each part of the matrix may provide better guidelines (e.g. clear numerical thresholds) on what constitutes high-performance versus something other than high-performance. Improvements in research and dissemination of findings could begin to address the research and industry needs in this area and potentially lead to greater effectiveness in AF deployment.

FUTURE RESEARCH

The potential impact of future research in this area is vast. There are several points of departure and several needs for additional research contained in this dissertation. Chapters 2, 3, and 4 present future research relating to that chapter. The significant areas of future research are presented collectively in the following sections.

Chapter 2 of the dissertation explores a broad view of goal and evaluative alignment between federal, state, and regional programs. The most practical future research that could come from this is the continuation of this methodology and a time-series comparison of how the levels of alignment change, in what goal domains, and by whom. Additionally, this part of the research sets the stage for an interesting (but extraordinarily complicated) optimization problem to find what levels of alignment *should* look like. As some of the reviewers of this paper (presented in Chapter 2) pointed out, there is literature that supports the theory that alignment is more effective and literature that supports the theory that less alignment is more effective. I surmise that it is dependent on the topic (the goal domain in this case). However, future research in this area is needed to further explore the effectiveness of alignment.

Chapter 3 sets the stage for two, well defined areas of research. First, there is a distinct need for technological and policy advancement related to two specific areas:

1. the need for an *automated and anonymous* way for stations to report throughput and the number of unique users at AF refueling stations without losing competitive advantage; and
2. the need for an *automated and anonymous* way to collect and report VMT and travel use patterns from private vehicles.

Advancement (e.g. invention) of technology in these two fields and advancements towards policy that allow the implementation of these technologies will improve the data attainability for these metrics. In addition, advancement in these areas will create a wealth of information that can be used to better define how and where AFVs are used and how often public AF refueling stations are being used. This knowledge will allow practitioners to more strategically locate refueling stations in the future. Perhaps the most important, area of research need exposed by this study is the need for measuring and quantifying the impact of outreach and education efforts. Specifically, there is a need to systematically identify what tools (e.g. test-drive events, webinars, inclusion of AF related curriculum in schools, etc.) are the most impactful for conveying not only the need for implementing AFVs, but the utility that AFVs may provide over traditionally fueled vehicles. In addition, there is a need to relate the relative impact of these different methods of outreach to cost so that practitioners and policymakers can know where there next outreach dollar is most effectively spent.

The assessment of evaluative reliability related to metrics presented in Chapter 4 emphasizes the need for further research in consistently and systematically identifying projects that use advanced fuel types. The findings from this chapter show that there is not a common understanding of how GFD should relate to high-performing projects that utilize advanced fuel types. Furthermore, the findings also show that metrics such as number of stations or number of vehicles are the most ambiguous. Together these findings highlight the need for metrics that can identify transformational technologies (e.g. high-performing projects) for advanced fuel types as there are currently few if any that can. The importance and need for future research in this specific area cannot be overstated.

The framework presented in this dissertation could be used to create and iteratively evaluate future metrics that can better discern potentially transformational technologies and applications. For example, Rogers (2003) describes a few factors of innovation in his now famous book “Diffusion of Innovation” as “relative advantage” and “compatibility” and “observability” to name a few. Imagine if we defined metrics related to compatibility as a level of cross-vehicle platform compatibility for a given fuel-type or project (e.g. # of combined vehicle refueling applications). Alternatively, imagine a metric related to Roger’s “observability” loosely defined as the degree to which “an innovation is visible to others” (Rogers, 2003) to address the outreach/education measurement gap discovered in this research. Taking these metrics through the second and third filters of the framework will allow the systematic vetting of these metrics. For example, in the second filter (Chapter 3) these metrics would be vetted for appropriateness, clarity, objectivity, and feasibility by a population of practitioners that will eventually use these metrics. This process will provide a fundamental and coherent definition of the metric. In addition, any problems associated with attaining the data would be uncovered *prior to implementation*. Metrics for that passed this filter would then pass to the third filter (Chapter 4) where a repeatable and reproducible definition of what high-performance means for each metric is established. Again, the use of this process provides practitioners with a better understanding of what metrics mean *before* we use them to gauge the relative progress and effectiveness of programs or projects.

CHAPTER REFERENCES

Almeida, C., 2011. Clean Vehicle Rebate Project – Final Report (FY 2009-2011). California Center for Sustainable Energy, Report G09-AQIP-02.

Alternative Fuels Data Center (AFDC), 2013. All Laws and Incentives Sorted by Type. United States Department of Energy, Alternative Fuels Data Center (online), retrieved June 10, 2013 from http://www.afdc.energy.gov/laws/matrix/incentive.

Bamberger, M., 2000. The Evaluation of International Development Programs: A View from the Front. American Journal of Evaluation, 21(1), 95–102.

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

Bureau of Transportation Statistics (BTS), 2011. National Transportation Statistics 2011, United States Department of Transportation, BTS (online), retrieved July 7, 2011 from http://www.bts.gov/publications/national_transportation_statistics/pdf/entire.pdf.

Byrne, M.R., Polonsky, M.J., 2001. Impediments to Consumer Adoption of Sustainable Transportation – Alternative Fuel Vehicles. International Journal of Operations and Production Management, 21(12), 1521-1538.

Chun, Y.H., Rainey, H.G., 2005. Goal Ambiguity and Organizational Performance in U.S. Federal Agencies. Journal of Public Administration Research Theory, 15(1), 1-30.

Clewell, B.C., Campbell, P.B., 2008. Building Evaluation Capacity: Designing a Cross-Project Evaluation (Guide 1). The Urban Institute, January 2008, available online at http://www.urban.org/UploadedPDF/411651_guide1.pdf.

Deutch, J.M., 2011. An Energy Technology Corporation Will Improve the Federal Government's Efforts to Accelerate Energy Innovation, Brookings Institution, Discussion Paper 2011-05, May 2011.

Federal Highway Administration (FHWA) (2008A). *SAFETEA-LU 1808: CMAQ Evaluation and Assessment – Phase I Report*. United States Department of Transportation, Federal Highway Administration (online), retrieved March 3, 2010 from http://www.fhwa.dot.gov/environment/cmaqpgs/safetealu1808/safetealu1808.pdf.

Federal Highway Administration (FHWA) (2008B). *SAFETEA-LU 1808: CMAQ Evaluation and Assessment – Phase II Report*. United States Department of Transportation, Federal Highway Administration (online), retrieved March 3, 2010 from http://www.fhwa.dot.gov/environment/cmaqpgs/fhwahep09026/fhwahep09026.pdf.

Gallagher, K.S., Collantes, G., Holdren, J.P., Lee, H., Frosch, R., 2007. Policy Options for Reducing Oil Consumption and Greenhouse-Gas Emissions from the U.S. Transportation Sector, Energy Technology Innovation Policy, Discussion Paper, July 2007.

Hallowell, M., 2008. A Formal Model of Construction Safety and Health Risk Management. Ph.D. dissertation, Oregon State Univ., Corvallis, Oregon.

Hallowell, M.R., Gambatese, J.A., 2010. Qualitative Research: Application of the Delphi Method to CEM Research. Journal of Construction Engineering and Management, 136(1), 99-107.

Heimann, L., 1995. Different Paths to Success: A Theory of Organizational Decision Making and Administrative Reliability. Journal of Public Administration Research and Theory, 5(1), 45-72.

Howell, D. and Chelius, M., 1997. Alternative Fuel Vehicle Programs – Applicability of Government Incentives. Transportation Research Record, Volume 1587, 1-9.

Hurst, E.G., 1980. Attributes of Performance Measures. Public Productivity Review, 4(1), 43-49.

International Energy Agency (IEA), 2002. Deployment Strategies for Hybrid, Electric, and Alternative Fuel Vehicles, Annex VIII Final Report.

International Energy Agency (IEA), 2011. Technology Roadmap – Electric and Plug-In Hybrid Electric Vehicles, Updated June 2011.

Johnson, C., 2012. Clean Cities Annual Metric Report 2011. National Renewable Energy Laboratories (NREL), Report NREL TP-7A30-56091.

Johnson, C., Bergeron, P., 2009. Clean Cities Annual Metric Report 2008. National Renewable Energy Laboratories (NREL), Report NREL TP-540-46424.

Kaplan, R.S., Norton, D.P., 1992. The Balanced Scorecard – Measures that Drive Performance. Harvard Business Review, Jan/Feb 1992, 1-10.

Kennerley, M., Neely, A., 2002. A Framework of the Factors Affecting the Evolution of Performance Measurement Systems. International Journal of Operations and Production Management, 22(11), 1222-1245.

Lee, H., 2009. Oil Security and the Transportation Sector, in: Gallagher, K.S. (Ed.), Acting in Time on Energy Policy. Brookings Institution Press, Washington, D.C.

Melaina, M., McQueen, S., Brinch, J., 2008. Refueling Infrastructure for Alternative Fuel Vehicles: Lessons Learned for Hydrogen – Workshop Proceedings, National Renewable Energy Laboratory (NREL), Proceedings NREL/BK-560-43669, July 2008.

Melendez, M., 2006. Transitioning to a Hydrogen Future: Learning from the Alternative Fuels Experience, National Renewable Energy Laboratory (NREL), Technical Report NREL/TP-540-39423, February 2006.

National Academy of Sciences (NAS), 2005. Thinking Strategically: The Appropriate Use of Metrics for the Climate Change Science Program, Committee on Metrics for Global Change Research, Climate Research Committee, National Research Council, National Academies Press, Washington, D.C.

National Renewable Energy Laboratory (NREL), 2011. Commercialization and Deployment at NREL – Advancing Renewable Energy and Energy Efficiency at Speed and Scale, NREL/MP-6A42-51947, May 2011.

Neely, A., Gregory, M., Platts, K., 1995. Performance Measurement System Design: A Literature Review and Research Agenda. International Journal of Operations & Production Management, 15(4), 80 – 116.

Pielke, R., 2009. The Honest Broker: Making Sense of Science in Politics and Policy, Cambridge University Press, Cambridge, United Kingdom.

Rogers, E., 2003. Diffusion of Innovations, 5th ed., Free Press, New York, New York.

Sagar, A., Gallagher, K.S., 2004. Energy Technology Demonstration and Deployment, in: National Commission on Energy Policy (NCEP) Technical Appendix for Chapter 6: Developing Better Energy Technologies for the Future, in: Ending the Energy Stalemate – A Bipartisan Strategy to Meet America's Energy Challenges, December 2004.

Schaumburg, Müller, 1996. Evaluating Capacity Building: Donor Support and Experiences. Report for the DAC (Development Assistance Committee) Expert Group on Aid Evaluation, OECD (Organization for Economic Cooperation and Development). Copenhagen, Denmark: DANIDA (Danish International Development Assistance), 1996.

Smith, P., 1990. The Use of Performance Indicators in the Public Sector. Journal of the Royal Statistical Society, 153(1), 53-72.

Sobin, N., Molenaar, K., Cahill, E., 2012. Mapping goal alignment of deployment programs for alternative fuel technologies: An analysis of wide-scope grant programs in the United States. Energy Policy, 51, 405-416.

Sperling, D., Gordon, D., 2009. Two Billion Cars: Driving Towards Sustainability, Oxford University Press, New York, New York.

Sperling, D., Yeh, S., 2010. Toward a Global Low Carbon Fuel Standard. Transport Policy, 17, 47-49.

Sperling, D., 1988. New Transportation Fuels – A Strategic Approach to Technological Change, University of California Press, Berkeley, California.

Stockdill, S.H., Baizerman, M., Compton, D.W., 2002. Toward a Definition of the ECB process: A Conversation with the ECB literature, in: Compton , D.W., Baizerman, M., Stockdill, S., (eds.), The Art, Craft, and Science of Evaluation Capacity Building, Vol. 93, New Directions for Evaluation, Jossey-Bass, San Francisco, CA.

Tesla Motors, Inc., 2013. Supercharger: The fastest recharging station on the planet, Tesla Motors, Inc., (online), June 10, 2013 from http://www.teslamotors.com/supercharger.

United States Energy Information Administration (USEIA), 2013. How many alternative fuel and hybrid vehicles are there in the U.S.?, USEIA (online), June 10, 2013 from http://www.eia.gov/tools/faqs/faq.cfm?id=93&t=4.

United States Government Accountability Office (USGAO), 2011. Opportunities to Reduce Potential Duplication in Government Programs, Save Tax Dollars, and Enhance Revenue. Report GAO 11-318SP, March 2011.

INTEGRATED REFERENCES

Almeida, C., 2011. Clean Vehicle Rebate Project – Final Report (FY 2009-2011). California Center for Sustainable Energy, Report G09-AQIP-02.
Alternative Fuels Data Center (AFDC), 2010. Data, Analysis, and Trends. United States Department of Energy, Alternative Fuels Data Center (online), retrieved February 17, 2010 from http://www.afdc.energy.gov/afdc/data/vehicles.html.
Alternative Fuels Data Center (AFDC), 2011. Federal & State Incentives & Laws. United States Department of Energy, Alternative Fuels Data Center (online), retrieved July 3, 2011 from http://www.afdc.energy.gov/afdc/laws/.
Alternative Fuels Data Center (AFDC), 2013. All Laws and Incentives Sorted by Type. United States Department of Energy, Alternative Fuels Data Center (online), retrieved April 13, 2013 from http://www.afdc.energy.gov/laws/matrix/incentive.
Alternative Fuels Data Center (AFDC), 2013. All Laws and Incentives Sorted by Type. United States Department of Energy, Alternative Fuels Data Center (online), retrieved June 14, 2013 from http://www.afdc.energy.gov/laws/matrix/incentive.
Bamberger, M., 2000. The Evaluation of International Development Programs: A View from the Front. American Journal of Evaluation, 21(1), 95–102.
Bang, G., 2010. Energy Security and Climate Change Concerns: Triggers for Energy Policy Change in the United States? Energy Policy, 38(4), 1645-1653.
Banzhaf, H.S., Chupp, B.A., 2010. Heterogeneous Harm vs. Spatial Spillovers: Environmental Federalism and US Air Pollution. National Bureau of Economic Research (NBER), Working Paper Series, NBER Working Paper 15666.
Behn, R.D., 2003. Why Measure Performance? Different Purposes Require Different Measures. Public Administration Review, 63(5), 586-606.
Brinkman, N., Wang, M., Weber, T., Darlington, T., 2005. Well-to-Wheels Analysis of Advanced Fuel/Vehicle Systems — A North American Study of Energy Use, Greenhouse Gas Emissions, and Criteria Pollutant Emissions, General Motors Corporation, Argonne National Laboratories, and Air Improvement Resource, Inc., Phase 2 Report May 2005.
Bureau of Transportation Statistics (BTS), 2011. National Transportation Statistics 2011, United States Department of Transportation, BTS (online), retrieved July 7, 2011 from http://www.bts.gov/publications/national_transportation_statistics/pdf/entire.pdf.
Burns, T., Stalker, G.M., 1961. The Management of Innovation, Tavistock, London, UK.
Byrne, M.R., Polonsky, M.J., 2001. Impediments to Consumer Adoption of Sustainable Transportation – Alternative Fuel Vehicles. International Journal of Operations and Production Management, 21(12), 1521-1538.
Calciolari, S., Cantu, E., Fattore, G., 2011. Performance Management and Goal Ambiguity: Managerial Implications in a Single Payer System. Health Care Management Review, 36(2), 164-174.
Chester, L., 2010. Conceptualising Energy Security and Making Explicit its Polysemic Nature. Energy Policy, 38(2), 887-895.
Chun, Y.H., Rainey, H.G., 2005. Goal Ambiguity and Organizational Performance in U.S. Federal Agencies. Journal of Public Administration Research Theory, 15(1), 1-30.
Churchman, C.W., 1968. The Systems Approach, Dell Publishing Co., New York, New York.
Clewell, B.C., Campbell, P.B., 2008. Building Evaluation Capacity: Designing a Cross-Project Evaluation (Guide 1). The Urban Institute, January 2008, available online at http://www.urban.org/UploadedPDF/411651_guide1.pdf.

Deutch, J.M., 2011. An Energy Technology Corporation Will Improve the Federal Government's Efforts to Accelerate Energy Innovation, Brookings Institution, Discussion Paper 2011-05, May 2011.

Dunlap, W.P., Burke, M.J., Smith-Crowe, K., 2003. Accurate Tests of Statistical Significance for rwg and Average Deviation and Interrater Agreement Indexes. Journal of Applied Psychology, 88(2), 356-362.

Federal Highway Administration (FHWA) (2008A). *SAFETEA-LU 1808: CMAQ Evaluation and Assessment – Phase I Report*. United States Department of Transportation, Federal Highway Administration (online), retrieved March 3, 2010 from http://www.fhwa.dot.gov/environment/cmaqpgs/safetealu1808/safetealu1808.pdf.

Federal Highway Administration (FHWA) (2008B). *SAFETEA-LU 1808: CMAQ Evaluation and Assessment – Phase II Report*. United States Department of Transportation, Federal Highway Administration (online), retrieved March 3, 2010 from http://www.fhwa.dot.gov/environment/cmaqpgs/fhwahep09026/fhwahep09026.pdf.

Finn, F.H., 1970. A Note on Estimating the Reliability of Categorical Data. Education and Psychological Measurement, 30, 71-76.

Gallagher, K.S., Collantes, G., Holdren, J.P., Lee, H., Frosch, R., 2007. Policy Options for Reducing Oil Consumption and Greenhouse-Gas Emissions from the U.S. Transportation Sector, Energy Technology Innovation Policy, Discussion Paper, July 2007.

General Motors Corporation (GMC), Argonne National Laboratory (ANL), BP, ExxonMobil, and Shell, 2001. Well-to-Wheel Energy Use and Greenhouse Gas Emissions of Advanced Fuel/Vehicle Systems — North American Analysis, Volumes 1-3, available online at http://greet.es.anl.gov/publications.

Greene, D.L., 2010. Measuring Energy Security: Can the United States Achieve Oil Independence? Energy Policy, 38(4), 1614-1621.

Gupta, J., Olsthoorn, X., Rotenberg, E., 2003. The Role of Scientific Uncertainty in Compliance with the Kyoto Protocol to the Climate Change Convention. Environmental Science & Policy 6(6), 475–486.

Gwet, K.L., 2010. Handbook of Inter-Rater Reliability, 2nd Ed., Advanced Analytics, LLC, Gaithersburg, Maryland.

Hallowell, M., 2008. A Formal Model of Construction Safety and Health Risk Management. Ph.D. dissertation, Oregon State Univ., Corvallis, Oregon.

Hallowell, M.R., Gambatese, J.A., 2010. Qualitative Research: Application of the Delphi Method to CEM Research. Journal of Construction Engineering and Management, 136(1), 99-107.

Heimann, L., 1995. Different Paths to Success: A Theory of Organizational Decision Making and Administrative Reliability. Journal of Public Administration Research and Theory, 5(1), 45-72.

Higgins, K.L., Maciariello, J.A., 2004. Leading Complex Collaboration in Network Organizations: A Multi-Disciplinary Approach, in: Beyerlein, M.M., Johnson, D.A., Beyerlein, S.T., (Eds.), Complex Collaboration: Building the Capabilities for Working Across Boundaries, Advances in Interdisciplinary Studies of Work Teams, Vol. 10, Elsevier, Ltd., Oxford, UK.

Holsti, O.R., 1969. Content Analysis for the Social Sciences and Humanities, Addison-Wesley Publishing, Reading, Massachusetts.

Howell, D. and Chelius, M., 1997. Alternative Fuel Vehicle Programs – Applicability of Government Incentives. Transportation Research Record, Volume 1587, 1-9.
Hurst, E.G., 1980. Attributes of Performance Measures. Public Productivity Review, 4(1), 43-49.
International Energy Agency (IEA), 2002. Deployment Strategies for Hybrid, Electric, and Alternative Fuel Vehicles, Annex VIII Final Report.
International Energy Agency (IEA), 2011. Technology Roadmap – Electric and Plug-In Hybrid Electric Vehicles, Updated June 2011.
James, L.R, Demaree, R.G., Wolf, G., 1984. Estimating Within-Group Interrater Reliability With and Without Response Bias. Journal of Applied Psychology, 69(1), 85-98.
Jansen, J.C., Seebregts, A.J., 2010. Long Term Energy Services Security: What is it and how can it be measured and valued. Energy Policy, 38(4), 1654-1664.
Jeon, C.M., Amekudzi, A.A., Guensler, R.L., 2013. Sustainability Assessment at the Transportation Planning Level: Performance Measures and Indexes. Transport Policy, 25, 10-21.
Johnson, C., 2012. Clean Cities Annual Metric Report 2011. National Renewable Energy Laboratories (NREL), Report NREL TP-7A30-56091.
Johnson, C., Bergeron, P., 2009. Clean Cities Annual Metric Report 2008. National Renewable Energy Laboratories (NREL), Report NREL TP-540-46424.
Jung, C.S., 2011. Developing and Validating New Concepts and Measures of Program Goal Ambiguity in the U.S. Federal Government. Administration and Society, (online before print), 1-27.
Kaplan, R.S., Norton, D.P., 1992. The Balanced Scorecard – Measures that Drive Performance. Harvard Business Review, Jan/Feb 1992, 1-10.
Kennerley, M., Neely, A., 2002. A Framework of the Factors Affecting the Evolution of Performance Measurement Systems. International Journal of Operations and Production Management, 22(11), 1222-1245.
Kruyt, B., van Vuuren, D.P., de Vries, H.J.M., Groenenberg, H., 2009. Indicators for Energy Security. Energy Policy, 37 (6), 2166–2181.
Lance, C.E., Butts, M.M., Michels, L.C., 2006. The Sources of Four Commonly Reported Cutoff Criteria: What did they really say? Organizational Research Methods, 9(2), 202-220.
Latham, G.P., Locke, E.A., 1991. Self-regulation through goal setting. Organizational Behavior and Human Decision Processes, 50, 212-247.
Lawrence, P.R., Lorsch, J.W., 1967. Organization and Environment: Managing Differentiation and Integration. Harvard University Press, Boston, MA.
LeBreton, J.M., Senter, J.L., 2008. Answers to 20 Questions About Interrater Reliability and Interrater Agreement. Organizational Research Methods, 11(4), 815-852.
Lee, H., 2009. Oil Security and the Transportation Sector, in: Gallagher, K.S. (Ed.), Acting in Time on Energy Policy. Brookings Institution Press, Washington, D.C.
Locke, E.A., 1968. Toward A Theory of Task Motivation and Incentives. Organizational Behavior and Human Performance, 3, 157-189.
Loschel, A., Moslener, U., Rubbelke, D., 2010. Indicators of Energy Security on Industrialised Countries. Energy Policy, 38(4), 1665-1671.
Lutsey, N., Sperling, D., 2008. America’s Bottom-Up Climate Change Mitigation Policy. Energy Policy, 36(2), 673-685.

Matland, R. E., 1995. Synthesizing the Implementation Literature: The Ambiguity-Conflict Model of Policy Implementation. Journal of Public Administration Research and Theory, 5, 145-174.

May, P., Saptotichne, J., Workman, S., 2006. Policy Coherence and Policy Domains. Policy Studies Journal, 34(3), 381-403.

Mazmanian, D., Sabatier, P., 1983. Implementation and Public Policy, Scott Foresman, Glenview, Illinois.

Melaina, M., McQueen, S., Brinch, J., 2008. Refueling Infrastructure for Alternative Fuel Vehicles: Lessons Learned for Hydrogen – Workshop Proceedings, National Renewable Energy Laboratory (NREL), Proceedings NREL/BK-560-43669, July 2008.

Melendez, M., 2006. Transitioning to a Hydrogen Future: Learning from the Alternative Fuels Experience, National Renewable Energy Laboratory (NREL), Technical Report NREL/TP-540-39423, February 2006.

Miles, R.E., Snow, C.C., Meyer, A.D., Coleman, H.J., 1978. Organizational Strategy, Structure, and Process. The Academy of Management Review, 3(3), pp.546-562.

Miller, D., Friesen, P., 1984. A Longitudinal Study of the Corporate Life Cycle. Management Science, 30, 1161-1183.

Mintzberg, H., 1979. The Structuring of Organizations, Prentice-Hall, Englewood Cliffs, NJ.

Mood, A., Graybill, F., Boes, D., 1974. Introduction to the Theory of Statistics, McGraw-Hill Book Company Inc., New York, New York.

Moynihan, D.P., 2008. The Dynamics of Performance Management: Constructing Information and Reform, Georgetown University Press, Washington, DC.

Nahorski, Z., Horabik, J., (2008). Greenhouse Gas Emission Permit Trading with Different Uncertainties in Emission Sources. Journal of Energy Engineering 134(2), 47–52.

National Academy of Sciences (NAS), 2005. Thinking Strategically: The Appropriate Use of Metrics for the Climate Change Science Program, Committee on Metrics for Global Change Research, Climate Research Committee, National Research Council, National Academies Press, Washington, D.C.

National Renewable Energy Laboratory (NREL), 2011. Commercialization and Deployment at NREL – Advancing Renewable Energy and Energy Efficiency at Speed and Scale, NREL/MP-6A42-51947, May 2011.

Neely, A., Gregory, M., Platts, K., 1995. Performance Measurement System Design: A Literature Review and Research Agenda. International Journal of Operations & Production Management, 15(4), 80 – 116.

Neuendorf, K.A., 2002. The Content Analysis Guidebook, Sage Publications, Thousand Oaks, California.

Ouchi, W.G., 1980. Markets, Bureaucracies, and Clans. Administrative Science Quarterly, 25(1), pp.129-141.

Pandey, S.K., Wright, B.E., 2006. Connecting the Dots in Public Management: Political Environment, Organizational Goal Ambiguity, and the Public Manager's Role Ambiguity. Journal of Public Administration Research Theory, 16(4), 511-532.

Parkinson, S., Begg, K., Bailey, P., Jackson, T., 2001. Accounting for Flexibility Against Uncertain Baselines: Lessons from Case Studies in the Eastern European Energy Sector. Climate Policy 1, 55–73.

Pielke, R., 2009. The Honest Broker: Making Sense of Science in Politics and Policy, Cambridge University Press, Cambridge, United Kingdom.

Plevin, R.J., 2010. Life Cycle Regulation of Transportation Fuels. Ph.D. dissertation. University of California at Berkeley, Berkeley, California.
Rainey, H.G., 2009. Understanding and Managing Public Organizations, 4th ed., Jossey-Bass, San Francisco, CA.
Ripley, R.B., Franklin, G.A., 1982. Bureaucracy and Policy Implementation. The Dorsey Press, Homewood, Illinois.
Rogers, E., 2003. Diffusion of Innovations, 5th ed., Free Press, New York, New York.
Saad. G.H., 2001. Strategic Performance Evaluation: Descriptive and Prescriptive Analysis. Industrial Management and Data Systems, 101(8), 390-399.
Sagar, A., Gallagher, K.S., 2004. Energy Technology Demonstration and Deployment, in: National Commission on Energy Policy (NCEP) Technical Appendix for Chapter 6: Developing Better Energy Technologies for the Future, in: Ending the Energy Stalemate – A Bipartisan Strategy to Meet America's Energy Challenges, December 2004.
Schaumburg, Müller, 1996. Evaluating Capacity Building: Donor Support and Experiences. Report for the DAC (Development Assistance Committee) Expert Group on Aid Evaluation, OECD (Organization for Economic Cooperation and Development). Copenhagen, Denmark: DANIDA (Danish International Development Assistance), 1996.
Sharma, C.H., 2011. Beyond Gaps and Imbalances: Restructuring the Debate on Intergovernmental Fiscal Relations. Public Administration, 90(1), 99-128.
Shen, Y., Lin, G., Li, K., Yuan, B., 2010. An Assessment of Exploitable Renewable Energy Sources with Concerns of Policy and Technology. Energy Policy, 38(8), 4604-4616.
Smith, P., 1990. The Use of Performance Indicators in the Public Sector. Journal of the Royal Statistical Society, 153(1), 53-72.
Sobin, N., Molenaar, K., Cahill, E., 2012. Mapping goal alignment of deployment programs for alternative fuel technologies: An analysis of wide-scope grant programs in the United States. Energy Policy, 51, 405-416.
Sovacool, B.K., 2011. An International Comparison of Four Polycentric Approaches to Climate and Energy Governance. Energy Policy, 39(6), 3832-3844.
Sperling, D., Yeh, S., 2010. Toward a Global Low Carbon Fuel Standard. Transport Policy, 17, 47-49.
Sperling, D., Gordon, D., 2009. Two Billion Cars: Driving Towards Sustainability, Oxford University Press, New York, New York.
Sperling, D., Yeh, S., 2009. Low Carbon Fuel Standards. Issues in Science and Technology (online), Winter 2009, retrieved June 16, 2011 from http://www.issues.org/25.2/sperling.html.
Sperling, D., 1988. New Transportation Fuels – A Strategic Approach to Technological Change, University of California Press, Berkeley, California.
Stockdill, S.H., Baizerman, M., Compton, D.W., 2002. Toward a Definition of the ECB process: A Conversation with the ECB literature, in: Compton , D.W., Baizerman, M., Stockdill, S., (eds.), The Art, Craft, and Science of Evaluation Capacity Building, Vol. 93, New Directions for Evaluation, Jossey-Bass, San Francisco, CA.
Tesla Motors, Inc., 2013. Supercharger: The fastest recharging station on the planet, Tesla Motors, Inc., (online), June 10, 2013 from http://www.teslamotors.com/supercharger.
United States Congress (USC), 1992. Energy Policy Act of 1992, Public Law 102-486, 102nd Congress.

United States Congress (USC), 1993. Government Results and Performance Act, Public Law 103-62, 103rd Congress.

United States Congress (USC), 2009. American Recovery and Reinvestment Act, Public Law 111-5, 111th Congress.

United States Department of Energy (USDOE), 2011. Database of State Incentives for Renewables and Efficiency (DSIRE). United States Department of Energy, Energy Efficiency and Renewable Energy (online), retrieved July 3, 2011 from http://www.dsireusa.org/.

United States Department of Energy (USDOE), 2011. Federal Energy Management Program: Laws and Regulations (online), retrieved November 7, 2011 from http://www1.eere.energy.gov/femp/regulations/regulations.html.

United States Energy Information Administration (USEIA), 2011. Alternatives to Traditional Transportation Fuels 2009, Released April 2011, USEIA (online), retrieved July 7, 2011 from http://www.eia.gov/renewable/alternative_transport_vehicles/pdf/afv-atf2009.pdf.

United States Energy Information Administration (USEIA), 2013. How many alternative fuel and hybrid vehicles are there in the U.S.?, USEIA (online), retrieved June 14, 2013 from http://www.eia.gov/tools/faqs/faq.cfm?id=93&t=4.

United States Government Accountability Office (USGAO), 2011. Opportunities to Reduce Potential Duplication in Government Programs, Save Tax Dollars, and Enhance Revenue. Report GAO 11-318SP, March 2011.

Vogel, D., Toffel, M., Post, D., Aragon, N., 2010. Environmental Federalism in the European Union and the United States. Harvard Business School (HBS), Working Paper 10-085.

Wagner, S.M., Rau, C., Lindemann, E., 2010. Multiple Informant Methodology: A Critical Review and Recommendations. Sociological Methods and Research, 38(4), 582

Winzer, C., 2012. Conceptualizing Energy Security. Energy Policy, Vol. 46, 36-48.

Witt, L.A., 1998. Enhancing Organizational Goal Congruence: A Solution to Organizational Politics. Journal of Applied Psychology, 83(4), 666-674.

Woodward, J., 1965. Industrial Organization: Theory and Practice, Oxford University Press, London, UK.

Wu, Y., Wang, M., Sharer, P., Rousseau, A., 2006. Well-to-Wheels Results of Energy Use, Greenhouse Gas Emissions, and Criteria Air Pollutant Emissions of Selected Vehicle/Fuel Systems. SAE 2006 Transactions (Journal of Engines), Paper No. 2006-01-0377.