The Policy Studies Journal, Vol. 36, No. 1, 2008

Assessing the Performance of Voluntary Environmental Programs: Does Certification Matter?

Nicole Darnall and Stephen Sides

The promotion of voluntary environmental programs (VEPs) as alternative approaches to traditional environmental regulation has fueled numerous researchers to evaluate VEP performance. However, these studies have focused on assessing the environmental performance of a single VEP. As yet, we know little about the overall environmental benefits of these programs. Moreover, questions remain about whether VEPs designed with different monitoring regimes—related to whether programs are self-monitored or undergo external certification—vary in their ability to improve environmental conditions. Using meta-analysis methodology, this article evaluates the aggregated environmental outcomes of U.S. VEPs drawing on data from nine studies and over 30,000 firms. We show that collectively VEP participants do not improve their environmental performance over nonparticipants. Rather, nonparticipants improve their environmental performance by 7.7 percent more than VEP participants. Additionally, nonparticipants improve the environment 24 percent more than participants in self-monitored VEPs, whereas participants in International Standards Organization 14001 as a group *exhibit inconclusive environmental performance improvements.*

KEY WORDS: voluntary environmental programs, certification, self-monitoring, environmental performance, meta-analysis

Introduction

Over the past 20 years, considerable efforts have been expended by governments, industry, and third-party organizations to develop and promote organizational participation in voluntary environmental programs (VEPs). In general, VEPs set forth a specific rationale to identify and guide the pursuit of improved environmental performance, in turn offering participants benefits for doing so.

Scholarly research has sought to explore the rationale for participation in VEPs and for undertaking certain activities that go beyond environmental compliance requirements (e.g., Arora & Cason, 1996; Bansal & Hunter, 2003; Khanna & Anton, 2002). Other studies have begun to recognize that VEP design characteristics may influence their proper functioning (Darnall & Carmin, 2005; Potoski & Prakash, 2005b; Rivera & deLeon, 2004). These evaluations have found that certain VEP design deficiencies, specifically the absence of third-party oversight of performance

Published by Blackwell Publishing. Inc., 350 Main Street, Malden, MA 02148, USA, and 9600 Garsington Road, Oxford, OX4 2DQ.

monitoring, invite "free-ridership" on the part of some participants, and "shirking" of underlying program commitments. VEPs lacking sufficient oversight therefore have been questioned for their intent and their benefits bestowed on participants.

Related to the environmental performance of VEPs, previous scholarship has suggested that VEPs—even those without third-party monitoring—are a means for improving participants' environmental performance (e.g., Khanna & Damon, 1999). However, in other instances, scholarship has demonstrated that participation in VEPs lacking oversight does not improve environmental performance (e.g., King & Lenox, 2000; Rivera & deLeon, 2004; Rivera, deLeon, & Koerber, 2006). Numerous questions therefore remain first about the overall utility of VEPs in that, as yet, we know little about whether these programs collectively are meeting their environmental goals. Second, questions remain about whether or not third-party monitoring of participants' adherence to program goals improves VEP efficacy.

To address these research needs, this study integrates the results of previous VEP research using meta-analytic techniques to develop a broader quantitative perspective of VEP performance. This analysis offers two contributions to organizational research on voluntary environmental governance. First, previous VEP scholarship has been equivocal in establishing a link between VEP participation and subsequent environmental improvements. One reason is that prior research has focused generally on evaluating the efficacy of individual programs and so a broader view of the efficacy of VEP participation is not understood. This study builds on previous VEP research findings to establish a much-needed overall assessment of their efficacy. In so doing, this research evaluates the entire body of peer-reviewed VEP literature and provides critical insights regarding the relationship between participation and improved environmental performance.

The second contribution of our analysis is that it distinguishes between the two most common subcategories of programs—those requiring self-monitoring as opposed to third-party certification. In making this distinction, we examine the proposition that International Organization for Standardization (ISO) 14001, which requires certification and independent third-party monitoring, improves participants' environmental performance to a greater degree than programs requiring self-monitoring alone. It offers evidence that *overall* VEP participants' environmental performance is worse than nonparticipants, especially for programs that require participants to self-monitor their adherence to program goals. Furthermore, participants in ISO 14001 *as a group* exhibit inconclusive environmental performance improvements.

Understanding VEPs

VEPs have been characterized under the broader rubric of voluntary environmental action, comprising several distinct categories of activity including negotiated agreements, unilateral projects, and sector and cross-sector alliances (Rondinelli & London, 2003). More formally, VEPs are defined as programs, codes, agreements, and commitments that encourage organizations to voluntarily reduce their environmental impacts beyond the requirements established by the environmental regulatory system

(Carmin, Darnall, & Mil-Homens, 2003). VEPs emerged in the late 1980s in response to criticisms about the overly prescriptive and often inefficient nature of the traditional command-and-control approach to environmental regulation (Davies, Mazurek, McCarthy, & Darnall, 1996). While there was widespread recognition of the problems with traditional regulation, there lacked consensus on how to achieve reform, particularly in the face of prevailing beliefs that restructuring would spur political and social confrontation (Davies et al., 1996). VEPs emerged as an alternative means for improving environmental conditions outside the regulatory development process (Carmin et al., 2003). Not only did these programs address a number of environmental concerns being raised by citizens and interest groups, they also avoided the complex and costly conflicts that often are associated with regulatory reform (Baggott, 1986).

Over 200 VEPs exist in the United States at the regional and national levels (Carmin et al., 2003), and even more programs operate within states and localities. The Environmental Protection Agency (EPA) is the largest sponsor of U.S.-based VEPs. By the end of the 1990s, some 13,000 companies were participating in EPA-sponsored VEPs (Mazurek, 2002). These programs account for 1.6 percent (Morgenstern & Pizer, 2007) of EPA's $4.3 billion operating budget, or approximately $69 million (USEPA, 2007). This cost has raised concerns from VEP critics who take issue with the fact that EPA has struggled to satisfy its Congressional mandate of ensuring environmental compliance within the regulated community. Agency officials and scholars alike argue that achieving regulatory compliance has lagged in large part because of limited Congressional funding for regulatory inspections and audits (Davies et al., 1996) and other budgetary reductions at a time when the agency's mandated responsibilities are increasing (Portney & Stavins, 2000). Given these circumstances, some suggest that government funding for VEPs should be reallocated toward helping ensure compliance with existing laws rather than managing nonregulatory programs, especially because empirical studies of VEP performance lack consistency.

In some instances, researchers have offered evidence that participation in government-sponsored VEPs improves environmental performance (e.g., Khanna & Damon, 1999) while other studies show that government-sponsored VEPs (e.g., Gamper-Rabindran, 2006) and VEPs that involved significant government involvement in their development (e.g., Rivera & deLeon, 2004; Rivera et al., 2006) do not. Studies that fail to show support for the idea that government-developed VEPs improve environmental performance attribute their findings to weak VEP design. VEPs having weak program structures are characterized by their inability to ensure that participants implement program goals, to require regular monitoring, or to mandate periodic reporting requirements (Darnall & Carmin, 2005). In instances where participant reporting is required, program managers rarely verify whether the reporting is accurate.

One reason why VEPs are developed with weak design structures is that program managers are balancing the need for rigor with the goal of providing a flexible means for participants to move beyond the parameters established by the traditional regulatory system. A tension therefore emerges between program managers' desires to encourage more widespread VEP participation and the need to ensure that program goals are met (Darnall, Carmin, Kreiser, & Mil-Homens, 2003). However, government agencies defend their VEPs. For instance, EPA claims that the general public has saved nearly $6 billion, conserved 603 million gallons of water, saved nearly 770 trillion British thermal units (BTUs) of energy and cut more than 438,000 tons of emissions (U.S. Office of Inspector General [USOIG], 2007).

Government VEPs are not the only programs under scrutiny in that concerns also have been raised about the utility of industry-sponsored VEPs. These programs generally focus on encouraging participation by firms within a specific industrial sector. The emergence of sector-specific industry VEPs was prompted by public concerns regarding the environmental performance of firms and by appeals to promote consistent environmental management among industry association members (Garcia-Johnson, 2000; Hoffman, 1997). Criticisms about these VEPs have been lodged by nongovernmental organizations (NGOs) and research scholars alike (e.g., Barber, 1998; King & Lenox, 2000) who have suggested that, like government-developed VEPs, VEPs designed entirely by industry associations lack appropriate implementation, monitoring, and reporting protocols that would lead to improved environmental performance outcomes.

One way to remedy VEP design weaknesses is to develop programs that require external third-party oversight and certification. Although less prevalent than other types of VEPs, certified VEPs have significant popularity because of their external monitoring and reporting features. ISO 14001 is the most widely recognized third-party-certified VEP. By April 2005, more than 88,800 facilities worldwide had certified to ISO 14001, of which, 4,671 facilities were U.S. based (Peglau, 2005). ISO 14001 requires participants to implement an environmental management system (EMS). EMSs consist of a collection of internal policies, assessments, plans, and implementation actions (Coglianese & Nash, 2001), affecting the entire organizational unit and its relationships with the natural environment. Once certified, the ISO 14001 label indicates that the company has implemented a management system that documents the firm's pollution aspects and impacts and identifies a pollution prevention process (Bansal & Hunter, 2003), which has continually improved over time (Darnall, 2006).

VEPs have several commonalities. First, they all require participant firms to establish environmental targets and statements of environmental goals (Darnall & Carmin, 2005). They also have comparable administrative requirements in the form of written agreements, letters of intent, or memoranda of understanding committing the participant to achieving program goals (Darnall & Carmin, 2005). However, these programs differ in one significant way. Government- and industry-sponsored VEPs generally are "self-monitoring" programs in that conformance with program goals is assured by VEP participants self-reporting their adherence to program requirements (Darnall & Carmin, 2005). By contrast, ISO 14001 requires external certification by an independent third party, which confers additional external legitimacy (Bansal & Hunter, 2003) about participants' environmental commitments. Furthermore, the certification process is more likely to formalize managerial commitment toward achieving environmental performance goals (Rondinelli & Vastag, 2000).

VEPs and Their Relationship with Environmental Performance

In spite of the equivocal findings in previous research, there are several reasons why we would expect all types of VEPs to improve participants' environmental performance. First, VEPs establish environmental performance goals, which may help focus participants' environmental management activities, enhance internal oversight of environmental improvement (King, Lenox, & Terlaak, 2005), and rally around a common cause. Second, most VEPs offer participants technical assistance, such as pollution prevention training, which can help firms meet program goals. Other ways in which VEPs can offer assistance is in the form of direct grants to hire environmental consultants. More prevalent in government-sponsored VEPs, direct grants subsidize VEP implementation increasing participants' access to external expertise that can fortify their environmental management capabilities (Darnall & Edwards, 2006). Subsidized technical assistance of all sorts, therefore, can encourage greater environmental performance gains.

Outside of creating environmental goals and providing technical assistance, VEPs generally establish peer networks among member firms. These networks allow for the exchange of information on best management practices. Information exchange also helps facilitate VEP implementation and may further assist participants in gaining knowledge that helps strengthen their environmental management capabilities. Furthermore, VEP participation can encourage greater collaboration within the firm. That is, by participating in a VEP, firms may improve their internal networks by collaborating across multiple operational areas, which creates broader organizational support for achieving VEP goals (Potoski & Prakash, 2005a). All these functional attributes of VEPs work collectively to encourage participants to improve their environmental performance.

Hypothesis 1: Organizations participating in VEPs have improved environmental performance.

VEP Oversight and Environmental Performance

Numerous VEP design features may affect participants' environmental performance, including program sponsorship. For instance, government-sponsored programs are designed with a greater diversity of stakeholder involvement, which avoids agency favoritism for one interest group or position (Carmin et al., 2003). At the same time, industry-sponsored programs are argued to lack rigorous implementation, monitoring, and reporting regimes and lead to lower environmental standards (Barber, 1998; King & Lenox, 2000). Despite these propositions, industry, government, and nonprofit VEP sponsors develop programs with similar administrative, environmental, and conformance requirements (Darnall & Carmin, 2005). As a result, other VEP design features, such as program oversight, may go further in explaining variations in environmental performance across different types of programs.

In general, there are three distinct "levels" of program oversight—internal assessments, self-reports to program sponsors, and external monitoring by an independent third-party auditor (Darnall & Carmin, 2005). Most program sponsors require participants to either self-report their environmental performance to program sponsors or rely on audits by independent third parties. For this reason, and because programs requiring internal assessments generally do not involve public disclosure of environmental information, previous empirical research typically has focused on the latter two types of VEPs to evaluate program efficacy.

In comparing VEPs that require self-reporting and external certification, we anticipate that certification programs may increase participants' environmental performance to a greater degree. This belief is based in large part on institutional arguments that external verification changes the dynamics of a collective action and causes individuals to more seriously consider their group obligations (Olson, 1965). In the case of VEPs, external certification may motivate participants to respond more earnestly to program requirements (Prakash, 2000) because participants undergo third-party evaluation that assesses their conformance to program standards. In the process, participants may be more likely to formalize managerial commitment toward achieving more ambitious environmental goals (Rondinelli & Vastag, 2000) and institutionalize existing pollution prevention programs by getting more value from what has already been developed (Darnall, 2006). Furthermore, independent review offers a greater degree of accountability and reduces opportunities for participants to behave opportunistically (Davies et al., 1996; King & Lenox, 2000). Additionally, companies that fail certification or have their certification revoked may face reputational penalties in the eyes of external stakeholders.

By contrast, VEPs that rely on self-reports of participants' environmental performance rarely undergo additional verification even by program sponsors (Darnall & Carmin, 2005). As such, these programs create opportunities for participants to report that they are achieving program goals when in fact they are not. In such instances, participants have an incentive to free ride and inaccurately "signal" that they are managing their environmental impacts to a greater extent than nonparticipants (Darnall & Carmin, 2005; Delmas & Keller, 2005). In so doing, participants may receive program benefits without changing their environmental behavior or meeting program goals. Because of the greater opportunities for free riding, we hypothesize that participants in self-monitored programs will improve their environmental performance to a lesser degree than participants in VEPs requiring external certification.

Hypothesis 2: Participation in certified VEPs is associated with greater improvements in environmental performance than participation in self-monitored VEPs.

Methods

To evaluate our hypotheses, we relied on meta-analytic techniques that considered the aggregate relationship between VEP participation and environmental performance. This approach was used to integrate the entire body of VEP literature to offer more generalized insights regarding the relationship between VEPs, their monitoring regimes, and their environmental performance.

Inclusion Criteria

The first step in preparing for the meta-analysis was to determine the population of studies relevant for inclusion. For purposes of this study, we relied on Carmin, Darnall, and Mil-Homens' (2003) definition of a VEP: any program, code, agreement, or commitment that encouraged business organizations to voluntarily reduce their environmental impacts beyond that required by the environmental regulatory system. The population of VEP studies was limited to programs that operate or operated in the United States (to control for the regulatory environment) and those that measured participants' environmental performance by way of pollution reductions. We defined environmental performance as an objective quantitative change in pollution or conditions contributing to the same (i.e., recycling, pollution prevention, and time-out of compliance). This definition focused the population of studies on those that demonstrated changes in actual or relative facility emissions, discharges and releases or off-site transfer of wastes, numbers of compliance citations, etc. Our definition excluded studies that did not consider pollution changes and eliminated studies that employed categorical or descriptive measures of environmental performance.

The population was further limited to studies evaluating regional or national programs and VEPs that were designed to have business organizations as their participants as opposed to government organizations, communities, or individuals. Suitable VEP studies were required to be published in a peer-reviewed journal. As such, they would have undergone extensive scrutiny by the scientific community.

Finally, there is an inevitable selection bias associated with VEP participation in that participation is voluntary. In the presence of selection bias, estimating the relationship between VEP participation and environmental performance would lead to an error term that is correlated with the participation decision. As the error term captures the effects of all omitted and imperfectly measured variables, any regressors that are correlated with the unmeasured or mismeasured factors will end up proxying for them. A two-stage estimation approach corrects for self-selection bias (Greene, 2003). Applied to the VEP setting, the factors that determine VEP participation (first stage) are estimated simultaneously with the factors that determine its environmental performance (second stage). To account for selection bias in the meta-analysis, the population of eligible studies was restricted to scholarship that utilized these two-stage estimation procedures.

Data

Studies that met our inclusion criteria were identified by conducting an exhaustive literature search using ABI/Inform® (ProQuest). ABI/Inform® provides access to the full text of over 18 hundred U.S. and international business, environmental, economic, and policy-related journals. Searches were conducted in May 2006 and covered the years 1982–2006. Cross-citations also were explored from previous VEP studies (i.e., Anton, Deltas, & Khanna, 2004; Darnall & Carmin, 2005; Delmas & Keller, 2005; Khanna & Anton, 2002; Lenox & Nash, 2003; Potoski & Prakash, 2005b; Rivera, 2002).

Some 60 studies evaluating aspects of VEPs were identified and evaluated for salient characteristics that would support their inclusion in the meta-analysis. Of these, 11 assessed the environmental performance of VEPs and controlled for selection bias. After further consideration, two of these studies were omitted. In one instance (Anton et al., 2004), the article evaluated adopters of an EMS without distinguishing whether adoption was part of a formal VEP (see Table 1). The second study (Rivera & deLeon, 2004) was excluded because an updated article (Rivera et al., 2006) was published in 2006 with more comprehensive data. The updated study is included in the meta-analysis.

Ultimately, nine studies met our selection criteria and were included in the analysis. This number is significant considering that a meta-analysis can be performed on as few as two qualifying studies because statistical power is limited only by the data utilized in the original studies (Lipsey & Wilson, 2001). As the total number of observations evaluated in the original studies is in excess of 30,000 firms, the meta-analysis had sufficient statistical power to arrive at a meaningful conclusion about the overall efficacy of VEPs that were included in this study.

Table 1. Voluntary Environment Program (VEP) Studies Excluded from the Meta-Analysis

Reference	VEP Evaluated	Dependent Variable	Empirical Approach and Exclusion Rationale
Anton, Deltas, & Khanna (2004)	Evaluates firms' EMS for evidence of increased comprehensiveness, TRI data, as well as separate assessment of regulatory and market pressures	TRI emissions, releases and off-site transfers, as well as hazardous air pollutant releases	Standard Poisson and negative binomial models assesses development of an EMS without regard to whether it was part of a VEP
Rivera & deLeon (2004)	Evaluates participants in the National Ski Areas Association's *Sustainable Slopes* program	Environmental "scorecard" based on quantitative measures	Two-stage probit model estimating the factors predicting both participation and environmental performance; excluded because an updated study was published in 2006 with more comprehensive data. The updated study is included in the meta-analysis.

Note: This table details studies that utilized some form of environmental performance measure and controlled for selection bias. Other VEP studies not listed were excluded because they did not evaluate environmental performance and instead focused on participation attributes or economic factors.
EMS, environmental management system.
TRI, Toxic Release Inventory.

Table 2. Summary of Voluntary Environment Programs (VEPs) Included in the Analysis

Name of Program	Brief Description
33/50 Program	• Sponsored by Environmental Protection Agency • Program that sought to engage participants in verifiable pollution prevention activities • One of the earliest government sponsored programs, attracted a large number of participants from U.S. industry • Goals were to reduce certain emissions, discharges and waste streams by 33% by 1992 and then 50% by 1995 • Participants self-reported their progress towards established goals
Climate Challenge Program	• Sponsored by the U.S. Department of Energy • Addresses an unregulated emission (carbon dioxide) • Participants encouraged to develop flexible approaches to achieve self-established goals • Formalized by written agreement
ISO 14001	• Sponsored by the International Organization for Standardization, a nonprofit organization, using input from numerous advisory groups and constituencies • Establishes an effective and verifiable environmental management system • Participation requires certification to the standard by an approved registrar • Administration of ISO 14001 is managed by registrars in each country • Participation expanded as a result of supply chain mandates
Responsible Care	• Sponsored initially by the Canadian Chemical Producers Association and later adopted by the U.S. Chemical Manufacturers Association (now known as the American Chemistry Council) • Required for all industry association members • Established codes for addressing critical health, safety and environmental management responsibilities • Participants self-reported their progress towards established goals • Only recently has the program required third-party performance audits
Sustainable Slopes Program	• Sponsored by the National Ski Areas Association • Involves 21 specific environmental “charters” for participating ski areas • Annual self-assessment required • Independent assessment of environmental performance used in analysis

ISO, International Organization for Standardization.

The nine studies evaluated 11 different environmental performance measures across five VEPs. Four VEPs required self-reports of environmental performance (33/50, Climate Challenge, Responsible Care, and Sustainable Slopes) and one required certification (ISO 14001). Table 2 describes each of these programs in greater detail. Table 3 offers a descriptive summary of the studies included in the meta-analysis.

Table 4 summarizes the data included in the meta-analysis calculations. It identifies the VEP type (self-monitoring or certified), the environmental performance measure used, as well as the effect size and weight for each environmental performance measure.

Empirics

There are three advantages to utilizing a meta-analytic approach. First, meta-analysis is less concerned with the underlying study statistics—whether they are significant or not—and more concerned with the relative “size” of the observed

Table 3. Voluntary Environment Program (VEP) Studies Included in the Meta-Analysis

Reference	VEP Evaluated	Dependent Variable	Empirical Approach
Gamper-Rabindran (2006)	Evaluates EPA's *33/50 Program* participants with respect to environ-mental performance	Percent reduction in 33/50 Program releases	Program participation integrated as a binary endogenous variable and compared to pre-33/50, 33/50, and post-33/50 program year releases. Two-stage model applied.
Khanna & Damon (1999)	Study of chemical industry participants in EPA's *33/50 Program* using TRI data	33/50 releases, Superfund site involvement	Two-stage least squares "participation" model, with predicted probability of participation and defined measure of environmental performance.
King & Lenox (2000)	Evaluates *Responsible Care* with respect to environmental performance and (estimated) facility production	Relative emissions and sector emissions using weighted TRI data	Covariance matrix and fixed-effect model relate VEP participation to environmental performance measures, while controlling for endogeneity through two-stage modeling.
King, Lenox, & Terlaak (2005)	Study of participants in *ISO 14001* utilizing TRI data	Calculated index of environmental performance focused on waste management	Discrete time, random-effect probit regression analysis two-stage) focusing on cross-sectional data to evaluate defined environmental performance measures.
Potoski & Prakash (2005a)	Analysis of U.S. facilities participating in *ISO 14001*	Logged and absolute index-weighted emissions	Treatment effects model and probit model, with endogeneity adjustments, focus on monitoring, and enforcement measures.
Potoski & Prakash (2005b)	Analysis of U.S. facilities participating in *ISO 14001*	Time-out of compliance with air pollution regulations	Two-stage OLS model of monitoring and enforcement measures affecting environmental performance.
Rivera, deLeon, & Koerber, (2006)	Follow-up study of Western ski resorts in the *Sustainable Slopes* program	Environmental "scorecard" based on quantitative measures	Recursive two-stage probit model using environmental performance as a predictor of participation.
Vidovic & Khanna (2007)	Evaluates EPA's *33/50 Program* participants with respect to environ-mental performance	Percent reduction in 33/50 Program releases	A two-step procedure involving a probit model predicting participation, followed by an analysis of participants' TRI releases.
Welch, Mazur, & Bretschneider (2000)	Evaluates data from largest electric utilities participating in EPA's *Climate Challenge Program*	Predicted and actual emissions for carbon dioxide and other priority air pollutants	Logit regression analysis with elements to address endogeneity. Beta coefficients provided on program participants representing high percentage of annual carbon dioxide emissions from electric utilities.

EPA, Environmental Protection Agency.
ISO, International Organization for Standardization.
TRI, Toxic Release Inventory.

Table 4. Description of Included Studies and Meta-Analytic Calculations

Author (Year)	Voluntary Environment Program (VEP) Name	Program Type	Environmental Performance Measure	Extracted Statistical Data[a]		Meta-Analysis Calculations	
				Difference Covariate Means	Regression Coefficient/ Result	Calculated Effect Size	Calculated Weight [1/(se)²]
Gamper-Rabindran (2006)	33/50	Self-monitored	33/50 releases—indexed to industry-type		0.54	0.54	22.72
Khanna & Damon (1999)	33/50	Self-monitored	33/50 releases—indexed to probability of participation	–2.40		–0.477	28.57
King & Lenox (2000)	Responsible Care	Self-monitored	(1) Relative emissions		0.08	–	–
			(2) Sector emissions		0.27	–	–
			Mean of (1) & (2)		0.175	0.36	117.6
Rivera, deLeon, & Koerber, (2006)	Sustainable Slopes	Self-monitored	Quantitative environmental "scorecard"	0.43		0.20	122
Vidovic & Khanna (2007)	33/50	Self-monitored	33/50 releases indexed to participation		0.109	0.88	59
Welch, Mazur, & Bretschneider (2000)	Climate Challenge	Self-monitored	Carbon dioxide reduction		0.09[b]	0.09	42
King, Lenox, & Terlaak (2005)	ISO 14001	Certified	(1) Index of Observed Waste Generation (Environmental Performance)		–0.02	–	–
			(2) Waste transfer off-site		0.03	–	–
			Mean of (1) & (2)		0.005	0.02	435
Potoski & Prakash (2005a)	ISO 14001	Certified	Logged Emissions Reduction—CERCLA-TRI Weighting		–0.0134	–0.03	109
Potoski & Prakash (2005b)	ISO 14001	Certified	Time-out of compliance with air pollution regulations		–0.0768	–0.16	142

[a]Unless otherwise noted, a negative sign denotes that the study found evidence that VEP participation was associated with reductions in pollution and a positive sign denotes that the study showed VEP participation was associated with increases in pollution.
[b]Study found no statistically significant relationship between VEP participation and pollution reductions.

effects. Second, meta-analysis corrects for sampling errors associated with studies of differing sizes (Hedges & Olkin, 1985). For instance, VEP studies with fewer observations are susceptible to sampling errors that in turn weaken the inferences that can be derived from that study. Meta-analysis corrects for these concerns by weighing the "effects" observed in individual studies according to the sample size, reducing the weight of small sample studies and, in turn, reducing sampling error in the comparison (Hunt, 1997). Third, meta-analysis allows for the quantitative comparison of multiple studies that evaluate related dependent variables (Lipsey & Wilson, 2001). These comparisons can be made regardless of whether the relationships in the original study are statistically significant (Hunter, 1997).

It is important to note that meta-analysis does not require that dependent variables be identical. Typically, there is variability among dependent variables because candidate studies inevitably have different measures of that variable (Glass, 1981; Hedges & Olkin, 1985; Hunter & Schmidt, 2004; Lipsey & Wilson, 2001). Furthermore, variability is anticipated among study features (including research models and data sources) and their attributes, and there is increasing acceptability to include more divergent study characteristics (Glass, 1981; Hedges & Olkin, 1985; Hunter & Schmidt, 2004; Lipsey & Wilson, 2001). For this reason, published meta-analytic research commonly utilizes broad measures of their dependent and independent variables (e.g., Boyle et al., 1994; Orlitzky & Benjamin, 2001; Ringquist, 2005). Related to the VEP setting, variability also exists in the population of studies evaluating environmental performance. Types of environmental performance measures that have been used include reductions in toxic releases (Khanna & Damon, 1999; King & Lenox, 2000; King et al., 2005; Potoski & Prakash, 2005a) and carbon dioxide emissions (Welch, Mazur, & Bretschneider, 2000), time-out of compliance with air pollution regulations (Potoski & Prakash, 2005b), and environmental scorecards that evaluate participant recycling and pollution prevention (Rivera et al., 2006).

With respect to accounting for how VEP participation was measured, all of the studies included in the analysis make use of either a dummy variable accounting for participation or utilize regression to establish a predicted probability of participation, which is then compared against actual participation to validate the empirical model. As such, actual comparison of the two groups—participants and nonparticipants—was possible.

While some studies included in the meta-analysis allowed for direct comparison of the covariate mean environmental performance of VEP participants and nonparticipants, others required that the effect size be estimated. To arrive at these estimations, we utilized regression coefficients (r) from previous studies to calculate an observed mean standardized effect level. This effect level characterizes the critical differences in environmental performance (a continuous variable) between VEP participants and nonparticipants (a dichotomous variable). It is calculated using the point-serial formula (Lipsey & Wilson, 2001):

$$ES = \delta = \frac{2r}{\sqrt{1-r^2}}$$

To correct for sampling and measurement error, studies were assigned a weight (ω) based on the inverse square of the standard error of the effect size (Lipsey & Wilson, 2001). This weight was calculated as a function of the number of VEP participants and nonparticipants in the study, and integrating a random effects constant (v) that accounts for variability across the observed effects:

$$\omega = \frac{1}{(se)^2 + v}, \quad \text{where} \quad se = \sqrt{\frac{n_1 + n_2}{n_1 n_2} + \frac{ES}{2(n_1 + n_2)}}$$

For some study correlation coefficients, the standard errors often are not provided or have a problematic determination, particularly where the r distribution is skewed and the population sample size is large (Cooper & Hedges, 1994). Given that the standard error is necessary in determining the sample weight, we used Lipsey and Wilson's (2001) procedure for calculating the Fisher Zr transformed correlation coefficient as an estimate of the effect size:

$$ES_{\mathrm{Zr}} = 0.5 \ln[1 + r/1 - r]$$

Using the Fisher Zr transformed correlation coefficients, the standard error calculation is as follows:

$$se = \sqrt{\frac{1}{n-3}}$$

Two of the studies included in our analysis used more than one environmental performance measure to assess the efficacy of a single VEP (i.e., King & Lenox, 2000; King et al., 2005). These studies had the potential to influence the meta-analysis results to a greater degree (Glass, McGaw, & Smith, 1981). To avoid bias related to the inclusion of more than one environmental performance measure per study, we took an *average* of the environmental performance measures and used this value in the meta-analysis. Doing so kept with conventions aimed at avoiding duplicate (covariate) contributions (Glass et al., 1981). The average weighted effect size across all studies in the meta-analysis was then calculated as follows (Lipsey & Wilson, 2001):

$$\overline{ES} = \frac{\sum (\omega \times ES)}{\sum \omega}$$

To assess whether or not the studies included in the analysis were estimating the same population mean, we performed a homogeneity test. This test assesses the variability in the data and defines the underlying model that can be used in calculating the effect size used for comparison purposes (Lipsey & Wilson, 2001). Homogeneity tests rely on a Q statistic, which is distributed chi-square. A positive homogeneity test would indicate similarity across the studies' measures and requires meta-analytic estimation by way of a fixed effects model. Otherwise, a random effects model should be used.

Table 5. Summary of Aggregate Findings—Relating Voluntary Environment Program (VEP) Participation to Environmental Performance

VEP Studies Included	Number of Environmental Performance Measures	Meta-Analysis Model	Meta-Analysis Outcome[a]	Standard Error	Z-Score	C.I. (95%)	Test of Heterogeneity[b]
All (9)	11	Random effects	0.077	0.030	2.6	0.018 to 0.136	Q = 69.0
Self-monitored (6)	7	Random effects	0.240	0.034	6.8	0.173 to 0.307	Q = 51.0
Certified (3)	4	Fixed effects	−0.025	0.038	0.7	−0.1 to 0.05	Q = 5.2*

[a]Meta-analysis outcome is an aggregate finding that combines the effects found in previous studies. A negative value indicates that VEP participation was associated with reductions in environmental impacts as compared to nonparticipants, whereas a positive value indicates that VEP participation was associated with increases in environmental impacts.
[b]A significant Q statistic indicates that data are homogeneous in nature and should be evaluated using a fixed effects model. An insignificant Q statistic indicates that data are heterogeneous and should be evaluated using a random effects model.
*Statistically significant at $p < 0.01$.

While our hypotheses differ from those in a typical empirical evaluation in that they are tested using results from previous studies, meta-analysis derives additional quantitative insights on the comparisons established in these original studies—and can support new findings. To evaluate Hypothesis 1, all studies were assessed without regard to their underlying monitoring regime. The results of our homogeneity test (Q = 69.0) indicated that a random effects model should be applied to evaluate the variability between overall VEP participation and environmental performance (see Table 5). To evaluate Hypothesis 2, the studies were divided into two distinct subsets—self-monitored programs and certified programs. Self-monitoring programs were defined as VEPs requiring participants to self-report their fulfillment of program expectations, whereas certified VEPs were defined as VEPs requiring independent verification that participants were fulfilling program requirements. Six studies evaluated self-monitoring programs (33/50, Climate Challenge, Responsible Care, and Sustainable Slopes). Based on the results of our homogeneity test (Q = 51.0), a random effects model was used to compare these programs. Three studies were recognized as evaluating certified VEPs and in each instance, those studies evaluated ISO 14001 participation.[1] Consistent with the fact that the three studies evaluated the same VEP, the results of our homogeneity test (Q = 5.2, $p < 0.01$) indicated that the data were homogeneous and that a fixed effects model should be applied to compare these programs.

While several of the studies evaluated the same VEP, each represents an independent assessment (sampling) of participants' environmental performance in addition to a unique dependent variable measure conforming to our definition. There are advantages to this approach in that having multiple studies of the same population and different measures of the dependent variable (Cook & Campbell, 1979; Cooper, 1989) allows for the correction for sampling error and improves the reliability of the relationship being explored (Webb, Campbell, Schwartz, Sechrest, & Grove, 1981).

Results

Table 5 describes the results of our meta-analysis models. It shows that across all programs, nonparticipants demonstrated a 7.7 percent stronger environmental performance than participants. Furthermore, the results of our confidence interval (95 percent) suggested that nonparticipants' environmental performance improved between 1.8 and 13.6 percent more than VEP participants. These findings did not support Hypothesis 1, which states that organizations participating in VEPs will have improved environmental performance.

When considering environmental performance in self-monitored programs, there was strong evidence that suggested participants failed to improve their environmental performance over nonparticipants. More specifically, nonparticipants demonstrated a 24 percent greater improvement in environmental performance than participants in self-monitored programs. Furthermore, a confidence interval (95 percent) of these findings indicated that the mean level of environmental performance improvement for nonparticipants was between 17 and 30 percent greater than participants in self-monitoring VEPs.

In evaluating the environmental performance of participants in ISO 14001, the results showed that participation was associated with a 2.5 percent improvement in environmental performance over nonparticipants. However, the related confidence interval (95 percent) indicated that the mean level of environmental performance improvements was between a 10 percent improvement and a 5 percent decrease in performance. As such, while overall performance may improve, because the confidence interval spans zero, there was inconclusive evidence regarding whether or not ISO 14001 certification leads to improved environmental performance.

Combined, these findings offer some support to Hypothesis 2, which states that participation in certified VEPs is associated with greater improvements in environmental performance than participation in self-monitored VEPs. However, they offer little evidence that participants in ISO 14001 had stronger environmental performance improvements than nonparticipants.

To consider the importance of VEP monitoring further, we assessed the variance explained in participants' environmental performance because of VEPs being either self-monitored or certified (ISO 14001). The Q statistics for the two groups were compared by way of a random effects analog to the one-way analysis of variance (ANOVA) (Lipsey & Wilson, 2001). This procedure offers equivalent results to estimating a regression model with a single dummy variable (self-monitored VEP participant or ISO 14001 certified participant). In so doing, the procedure compares estimates the "within" study variation and the "between" study variation to arrive at an overall conclusion as to whether the self-monitored and certified VEPs demonstrate equivalent environmental performance improvements. The results of the ANOVA analog show that the Q "between study" group data ($Q = 12.8$, $p < 0.01$) account for a significant amount of the variability in our data, which affirms that our

Table 6. Analysis of Variance (ANOVA) Analog Comparison of Voluntary Environment Program (VEP) Groupings

	Heterogeneity Measure—Q-Statistic				
	All VEPs (Q_{TOT})	Self-Monitored VEPs (Q_{SELF})	Certified VEPs (Q_{CERT})	Within Study Groups ($Q_W = Q_{CERT} + Q_{SELF}$)	Between Study Groups ($Q_B = Q_{TOT} - Q_W$)
Number of studies	9	6	3	–	–
Degrees of freedom	8	5	2	7 (9 studies – 2 groups)	1 (2 groups – 1)
Calculated Q—see Table 5	69.0	51.0	5.2	–	–
Calculated Q—ANOVA analog[a]	–	–	–	56.2	12.8*

[a]The ANOVA analog tests for whether or not groupings (related to self-monitored and certification programs) account for significant variability in observed environmental performance improvements.
*Statistically significant at $p < 0.01$, indicating that the monitoring grouping explains significant variability in environmental performance.

groupings are statistically valid (see Table 6). Differences in VEP monitoring regimes therefore explain significant variation in participants' environmental performance changes.

Discussion and Conclusions

Scholars and practitioners have presented numerous concerns about the overall utility of VEPs, especially because we know little about whether these programs collectively are meeting their environmental goals. This research helps address these issues by aggregating the results found in previous studies to arrive at an overall conclusion about the efficacy of VEPs. Our findings indicate that there is little evidence that *overall* VEP participation is associated with improved environmental performance. Rather, nonparticipants improve the environment 7.7 percent more than VEP participants. These unexpected findings are explained in part by differences in VEP design in that significant variations exist among the two most common subcategories of programs—those requiring self-monitoring as opposed to those requiring third-party certification.

In comparing these two VEP monitoring regimes, this study shows that, *on average*, nonparticipants demonstrate a 24 percent stronger environmental performance improvement than participants in self-monitored VEPs. By contrast, participants in ISO 14001, which requires external certification by an independent third-party auditor, *on average*, shows modest environmental performance improvements (2.5 percent) over nonparticipants. However, because the mean level of environmental improvements for ISO 14001 participants ranges between 10 and –5 percent, we are not confident that ISO 14001 certification leads to real environmental improvements.

Participants in self-monitored VEPs self-evaluate their adherence to program requirements and submit reports to program managers of their progress. However, conformance is rarely verified, and if it is, there exists no mechanism to expel participants that fail to meet the VEP's environmental goals (Darnall & Carmin, 2005). Situations such as these invite widespread free ridership and shirking of VEP goals (King & Lenox, 2000; Rivera, 2002). Participants, therefore, derive program benefits (such as increased recognition for a proactive environmental strategy) without conforming to VEP requirements (Darnall & Carmin, 2005).

While ISO 14001 requires certification by an independent third-party auditor, which reduces some opportunities for free riding, it fails to incorporate two important VEP design features that would help ensure that participants improve their environmental performance. First, ISO 14001 does not require that third-party audits be publicly disclosed. Such disclosure may succeed in improving participants' environmental performance because they are held accountable to a greater degree (Potoski & Prakash, 2005a). The second design weakness of ISO 14001 is that it does not have a strong sanctioning mechanism. Indeed, firms that are granted ISO 14001 certification rarely are decertified if they fail to maintain conformance with the standard (Barrow, 2005). Moreover, in the unlikely event that a firm is decertified, ISO does not publish summary statistics for these companies. This design policy differs from positions taken by other certified VEPs. For instance, 16 participants of the Sustainable Forestry Initiatives' (SFI) original 45 members were expelled from the program for failure to uphold the standard (American Forest and Paper Association, 2001). SFI publicly disclosed conformance information on its website to maintain external credibility for its VEP and enhance the legitimacy of participants that successfully maintain their certification.[2]

To address the problem of free riding, participants in both self-monitored and certified VEPs need to be separated in some way such that companies that fulfill program requirements are differentiated from participants that fail to do so. Imposing greater monitoring and expelling free riders is one way to create this separation, in that firms intending to free ride would be discouraged from participating in a VEP or removed for their nonconformance (Darnall & Carmin, 2005). By preventing shirking—in both sponsor-monitored and certified VEPs—VEPs would encourage a virtuous cycle of reputational improvements for VEPs and their participants. That is, participants would be more likely to contribute to maintaining the VEP's reputation because they believe other members will do so as well (Potoski & Prakash, 2005b; Scholz & Lubell, 1998).

However, weak VEP goals may explain why program participants do not demonstrate greater environmental improvements than nonparticipants. Goal setting theory suggests that goals serve an energizing function (Locke & Lantham, 2002) for organizations. Specific and challenging goals result in a higher performance than moderate or easy attainable goals (Locke & Latham, 1990). The highest level of effort generally occurs when a task is moderately difficult and the lowest levels occur when the task is either very easy or very hard (Atkinson, 1958). Relating these issues to VEPs, in fact participants may be adhering to program goals. However, if programs are designed with goals that are weaker than the internal environmental performance targets established by nonparticipating firms, participants will be less likely to outperform nonparticipants.

We originally hypothesized that participation in ISO 14001 would lead to greater environmental performance changes than participation in self-monitored programs.

While our findings support this proposition, it is only because self-monitored programs appear to perform so poorly. Indeed, both types of programs fail to offer compelling evidence that their participants improve the environment to a greater degree than nonparticipants. Combined, our findings point to the hazards of developing VEPs with weak design structures. While it is important to recognize that program managers are balancing the need for ensuring program efficacy with the goal of providing participants flexibility, flexibility in program design (as currently implemented) can compromise the ability of these programs to improve participants' environmental performance.

Our findings have important implications for policymakers who promote VEPs in that the programs evaluated in this study do not demonstrate any real environmental performance improvements. However, it is important to ask, "what is the role of these programs?" If VEPs are designed for the single purpose of encouraging participants to improve the environment to a greater degree than nonparticipants, then these programs may have little tangible merit. These conclusions are consistent with recent evaluations criticizing government-sponsored VEPs for their inability to determine success, failure, and ideal models for future program development (USOIG, 2007). In the absence of developing stronger design features, VEP critics may be justified in their concerns about the overall utility of these programs.

However, VEPs may also serve a more nuanced (but equally important) role. Even if a firm's participation represents a purely symbolic act, many VEP participants implemented performance-improving environmental practices prior to the existence of the VEP. Strong empirical evidence of this relationship has been found for ISO 14001 adopters (King, Lenox, & Terlaak, 2005) as well as companies that participated in EPA's 33/50 program (Khanna & Damon, 1999). That is, many firms use VEP participation as a vehicle for communicating their previous environmental improvements in an effort to gain external social and economic rewards (King et al., 2006). In instances such as these, there is a "decoupling" between substantive environmental improvements related to the symbolic act of VEP participation (King et al., 2006). Coupling of symbol and substance then occurs after the VEP is developed in that the program serves as a vehicle for firms to credibly communicate their prior activities (King et al., 2006). In the absence of the VEP, there would be no way for companies to communicate their previous environmental improvements and obtain value from them. However, prior environmental improvements may also increase the marginal cost for VEP participants to achieve greater pollution reductions because their low-hanging fruit has been picked. By contrast, to the extent that nonparticipating firms are engaging in pollution reductions for the first time, they can more readily make small changes in their operating procedures that lead to greater environmental gains. These distinctions may partially explain why VEP participants do not demonstrate greater environmental improvements.

It is also important to recognize that VEPs rely on flexible approaches to encourage environmental improvement. In some instances, this flexibility may help participants foster collaborative relationships between government and the regulated community that promotes shared learning and capacity development (Darnall & Carmin, 2005). In other instances, VEPs have influenced corporate attitudes regarding the environment and management practices. While these activities may have little impact on VEP efficacy, they can create a foundation for long-term environmental stewardship (Darnall & Carmin, 2005; Morgenstern & Pizer, 2007).

Related to the regulatory realm, there may be additional merit for promoting VEPs outside of demonstrating efficacy. For instance, VEPs can investigate and promote innovative environmental policy ideas (Delmas & Terlaak, 2001) when political resistance prevents the adoption of more powerful mandatory plans (Lyon, 2003). Such ideas can be tested and evaluated before they are implemented across the regulated community (Darnall & Carmin, 2005). As such, in asking the question "what is the role of these programs?" the answer may be much broader than was originally conceived by their program designers.

In arriving at our conclusions, this research imposed numerous inclusion and exclusion criteria before deeming a study suitable for comparison in this analysis. Despite our detailed approach, there are reasons to interpret our findings cautiously. The limited number of existing VEP studies prevented us from utilizing a multiple regression analysis. Rather, this study evaluated the relationship between VEP monitoring regimes and participants' environmental performance by way of a random effects analog to the one-way ANOVA. The ANOVA analog offers useful insights about the relationship between VEP participation, program monitoring, and participants' environmental performance (D. Wilson, personal communication, September 17, 2007). While our findings draw on more than 30,000 observations of VEP participant and nonparticipant behavior, and include VEPs that had reporting, monitoring, administrative, and conformance requirements, which are typical of many VEPs, we believe care should be exercised in extrapolating our findings to predict the environmental performance outcomes of other VEPs, especially those with more (or less) robust accountability requirements. Moreover, this study did not account for the important, but difficult to quantify intangible effects of VEP participation, such as changes in attitudes or management practices. While these issues may have a trivial influence on VEP efficacy, they can lead to behavioral changes that benefit the natural environment in the future (Darnall & Carmin, 2005; Morgenstern & Pizer, 2007).

As the variety and number of VEPs grows and as markets and other stakeholders increasingly rely on these programs to gauge firms' environmental performance, it is essential to understand more about their purpose and efficacy. Additionally, as more resources are dedicated to VEP development and implementation, it will be increasingly important to understand which design characteristics lead to greater environmental improvements. Future research should evaluate this latter issue in particular. For instance, some VEPs (e.g., Responsible Care) were designed as self-monitored VEPs[3] but recently added third-party auditing requirements. By studying programs such as these, we might be able to better understand the impact that monitoring regimes have on participant shirking and the extent to which third-party monitoring helps achieve program goals within the same VEP. Such an evaluation also would offer stronger empirical evidence regarding the merit of external auditing. Additionally, future studies would benefit by comparing the efficacy of ISO 14001 with other certified VEPs (like SFI) that couple certification with public disclosure requirements in addition to sanctions that expel nonconforming participants.

Doing so would offer important evidence about the incremental environmental performance gains offered by various VEP design features.

In sum, the research presented in this article builds on the inconsistent findings of previous research to establish a much-needed overall assessment of VEP efficacy. It offers two contributions to organizational research on voluntary environmental governance, in addition to areas for future study. First, it provides a broader view of the efficacy of VEPs as alternative environmental governance tools and presents critical insights regarding the relationship between participation and improved environmental performance. Second, this research distinguishes among the two most common subcategories of programs—those requiring self-monitoring as opposed to third-party certification. It offers evidence that *overall* VEP participants' environmental performance is worse than nonparticipants, especially for programs that require participants to self-monitor their adherence to program goals. Furthermore, participants in ISO 14001 *as a group* exhibit inconclusive environmental performance improvements. Finally, this research demonstrates the need for future research to consider the varying purposes of VEPs and how specific design features can increase their value as tools for environmental governance.

Nicole Darnall is Assistant Professor of environmental science and policy at George Mason University. Her research interests include environmental policy innovation and industry self-regulation. E-mail: ndarnall@gmu.edu.
Stephen Sides is a doctoral student of environmental science and policy at George Mason University. His research interests include industry self-regulation, valuation of environmental management systems, and public health regulatory reform. E-mail: ssides1@gmu.edu.

Notes

1. ISO 14000 series consists of 23 standards constituting the ISO 14000 "family." One such standard is ISO 14001. This standard *requires* certification, whereas the others do not. In some instances, a firm may claim it conforms to ISO 14001, but does not obtain certification. However, our analysis does not include studies that consider whether or not individual participants claim to be certified. Rather, it includes studies that examine actual certification.
2. SFI was developed as an industry-initiated third-party certification VEP. Industry representatives worked in concert with representatives from the environmental, professional, conservation, academic, and public sectors to design this program. On January 1, 2007, the program became fully independent and it is governed by SFI, Inc.
3. King and Lenox (2000) evaluated the efficacy of Responsible Care. Their study was completed during the time in which the program required self-monitoring only.
4. *Denotes study included in meta-analysis.

References[4]

American Forest and Paper Association. 2001. *The SFI Program: A Bold Approach to Sustainable Forest Management* [Online]. http://www.afandpa.org/forestry/sfi_frame. Accessed December 2001.

Anton, Wilma Rose Q., George Deltas, and Madhu Khanna. 2004. "Incentives for Environmental Self-Regulation and Implications for Environmental Performance." *Journal of Environmental Economics and Management* 48 (1): 632–54.

Arora, S., and T. N. Cason. 1996. "Why Do Firms Volunteer to Exceed Environmental Regulations? Understanding Participation in EPA's 33/50 Program." *Land Economics* 72: 413–32.

Baggott, Robert. 1986. "By Voluntary Agreement: The Politics of Instrument Selection." *Public Administration* 64 (1): 51–67.

Bansal, Pratima, and Trevor Hunter. 2003. "Strategic Explanations for the Early Adoption of ISO 14001." *Journal of Business Ethics* 46 (3): 289–99.

Barber, Jeffrey. 1998. "Responsible Action or Public Relations? NGO Perspectives on Voluntary Initiatives." *Industry and Environment* 21 (1–2): 19–22.

Barrow, Christopher J. 2005. *Environmental Management and Development*. London: Taylor & Francis, Inc.

Boyle Kevin J., Gregory L. Poe, and John C. Bergstrom. 1994. "What Do We Know about Groundwater Values? Preliminary Implications from a Meta Analysis of Contingent Valuation Studies." *American Journal of Agricultural Economics* 76 (December): 1055–61.

Carmin, JoAnn, Nicole Darnall, and Joao Mil-Homens. 2003. "Stakeholder Involvement in the Design of U.S. Voluntary Environmental Programs: Does Sponsorship Matter?" *Policy Studies Journal* 31: 527–43.

Coglianese, Cary, and Jennifer Nash, eds. 2001. *Regulating from the Inside: Can Environmental Management Systems Achieve Policy Goals?* Washington, DC: Resources for the Future.

Cook Thomas D., and Donald T. Campbell. 1979. *Quasi Experimentation: Design and Analysis Issues for Field Settings*. Boston: Houghton Mifflin.

Cooper, Harris M. 1989. *Integrating Research: A Guide for Literature Reviews*, 2nd ed. Newberry Park, CA: Sage Publications.

Cooper, Harris M., and L. V. Hedges. 1994. *The Handbook of Research Synthesis*. Newberry Park, CA: Sage Publications.

Darnall, Nicole. 2006. "Why Firms Mandate ISO 14001 Certification." *Business & Society* 45 (3): 354–81.

Darnall, Nicole, and JoAnn Carmin. 2005. "Cleaner and Greener? The Signaling Accuracy of U.S. Voluntary Environmental Programs." *Policy Sciences* 38: 71–90.

Darnall Nicole, JoAnn Carmin, Nicole Kreiser, and Joao Mil-Homens. 2003. *The Design & Rigor of U.S. Voluntary Environmental Programs: Results from the VEP Survey*. Department of Political Science & Public Administration, North Carolina State University and Department of Urban Studies and Planning, Massachusetts Institute of Technology.

Darnall Nicole, and Daniel Edwards, Jr. 2006. "Predicting the Cost of Environmental Management System Adoption: The Role of Capabilities, Resources and Ownership Structure." *Strategic Management Journal* 27 (2): 301–20.

Davies Terry, Janice Mazurek, Kieran McCarthy, and Nicole Darnall. 1996. *Industry Incentives for Environmental Improvement: Evaluation of U.S. Federal Initiatives*. Washington, DC: Resources for the Future, Center for Risk Management.

Delmas, Magli, and A. Keller. 2005. "Free Riding in Voluntary Environmental Programs: The Case of the U.S. EPA WasteWise Program." *Policy Sciences* 38 (2–3): 91–106.

Delmas, Magali A., and Ann K. Terlaak. 2001. "A Framework for Analyzing Environmental Voluntary Agreements." *California Management Review* 43: 44–62.

*Gamper-Rabindran, Shanti. 2006. "Did the EPA's Voluntary Industrial Toxics Program Reduce Emissions? A GIS Analysis of Distributional Impacts and By-Media Analysis of Substitution." *Journal of Environmental Economics and Management* 52: 391–410.

Garcia-Johnson, Ronnie. 2000. *Exporting Environmentalism: U.S. Multinational Chemical Corporations in Brazil and Mexico*. Cambridge, MA: MIT Press.

Glass, Gene V., Barry McGaw, and Mary L. Smith. 1981. *Meta-Analysis in Social Research*. Beverly Hills, CA: Sage Publications.

Greene, William H. 2003. *Econometric Analysis*, 5th ed. Upper Saddle, NJ: Prentice Hall.

Hedges, Larry, and Ingram Olkin. 1985. *Statistical Methods for Meta-Analysis*. San Diego, CA: Academic Press, Ltd.

Hoffman, Andrew. 1997. *From Heresy to Dogma: An Institutional History of Corporate Environmentalism*. San Francisco: New Lexington Press.

Hunt, Morton. 1997. *How Science Takes Stock: The Story of Meta-Analysis*. New York: Russell Sage Foundation.

Hunter, John E. 1997. "Needed: A Ban on the Significance Test." *Psychological Science* 8: 3–7.

Hunter, John E., and Frank Schmidt. 2004. *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*. Thousand Oaks, CA: Sage Publications.

Khanna, Madhu, and Wilma Rose Anton. 2002. "Corporate Environmental Management: Regulatory and Market-Based Incentives." *Land Economics* 78: 539–58.

*Khanna, Madhu, and L. Damon. 1999. "EPA's Voluntary 33/50 Program: Impact on Toxic Releases and Economic Performance of Firms." *Journal of Environmental Economics and Management* 37 (1): 1–17.

*King, Andrew A., and Michael Lenox. 2000. "Industry Self-Regulation Without Sanctions: The Chemical Industry's Responsible Care Program." *Academy of Management Journal* 43 (4): 698–716.

*King, Andrew A., Michael Lenox, and Ann Terlaak. 2005. "The Strategic Use of Decentralized Institutions; Exploring Certification with the ISO 14001 Management Standard." *Academy of Management Journal* 48 (6): 1091–106.

Lenox, Michael, and Jennifer Nash. 2003. "Industry Self-Regulation and Adverse Selection: A Comparison Across Four Trade Association Programs." *Business Strategy and the Environment* 12 (6): 343–60.

Lipsey, Mark W., and David B. Wilson. 2001. *Practical Meta-Analysis*. Thousand Oaks, CA: Sage Publications, Inc.

Locke, Edwin A., and Garry P. Latham. 2002. "Building a Practically Useful Theory of Goal Setting and Task Motivation." *American Psychologist* 57: 705–15.

———. 1900. *A Theory of Goal Setting and Task Performance*. Englewood Cliffs, NJ: Prentice Hall.

Lyon, Thomas P. 2003. *Voluntary versus Mandatory Approaches to Climate Change Mitigation*. Washington, DC: Resources for the Future, Issue Brief 03–1.

Mazurek, Janice. 2002. "Government Sponsored Voluntary Programs for Firms: An Initial Survey." In *New Tools for Environmental Protection: Education, Information and Voluntary Measures*, ed. T. Dietz and P. C. Stern. Washington, DC: National Academy Press, 219–34.

Morgenstern, Richard, and William Pizer. 2007. *Reality Check: The Nature and Performance of Voluntary Environmental Programs in the United States, Europe, and Japan*. Washington, DC: Resources for the Future.

Olson, Mancur. 1965. *The Logic of Collective Action*. Cambridge, MA: Harvard University Press.

Orlitzky, Marc, and John Benjamin. 2001. "Corporate Social Performance and Firm Risk: A Meta-Analytical Review." *Business and Society* 40 (4): 369–96.

Peglau, Reinhard. 2005. *ISO 14001 Certification of the World*. Berlin: Federal Environmental Agency.

Portney, Paul R., and Richard N. Stavins, eds. 2000. *Public Policies for Environmental Protection*. Washington, DC: Resources for the Future.

*Potoski, Matthew, and Aseem Prakash. 2005a. "Covenants with Weak Swords: ISO 14001 and Facilities' Environmental Performance." *Journal of Policy Analysis and Management* 24 (4): 745.

*———. 2005b. "Green Clubs and Voluntary Governance: ISO 14001 and Firms' Regulatory Compliance." *American Journal of Political Science* 49 (2): 235.

Prakash, Aseem. 2000. *Greening the Firm*. Cambridge, UK: Cambridge University Press.

Ringquist, Evan. 2005. "Assessing Evidence of Environmental Inequities: A Meta-Analysis." *Journal of Policy Analysis and Management* 24 (2): 223–47.

Rivera, Jorge. 2002. "Assessing a Voluntary Environmental Initiative in the Developing World: The Costa Rican Certification of Sustainable Tourism." *Policy Sciences* 35: 333–60.

Rivera, Jorge, and Peter deLeon. 2004. "Is Greener Whiter? Voluntary Environmental Performance of Western Ski Areas." *Policy Studies Journal* 32 (3): 417–38.

*Rivera, Jorge, Peter deLeon, and Charles Koerber. 2006. "Is Greener Whiter Yet? The Sustainable Slopes Program After Five Years." *Policy Studies Journal* 34 (2): 195–224.

Rondinelli, Dennis A., and Ted London. 2003. "How Corporations and Environmental Groups Cooperate: Assessing Cross Sector Alliances and Collaborations." *Academy of Management Executive* 17 (1): 62–76.

Rondinelli, Dennis A., and Gyula Vastag. 2000. "Panacea, Common Sense, or Just a Label? The Value of ISO 14001 Environmental Management Systems." *European Management Journal* 18: 499–510.

Scholz, John T., and Mark Lubell. 1998. "Adaptive Political Attitudes: Duty, Trust and Fear as Monitors of Tax Policy." *American Journal of Political Science* 42: 398–417.

U.S. Environmental Protection Agency. 2007. *FY2008 EPA Budget in a Brief*. Washington, DC: USEPA, Office of the Chief Financial Officer, EPA-205-S-07-001.

U.S. Office of Inspector General. 2007. *Voluntary Programs Could Benefit from Internal Policy Controls and Systematic Management Approach*. Report No. 2007-P-00041, September 25. Washington, DC: USOIG.

*Vidovic, Martina, and Neha Khanna. 2007. "Can Voluntary Pollution Prevention Programs Fulfill Their Promises? Further Evidence from the EPA's 33/50 Program." *Journal of Environmental Economics and Management* 53: 180–95.

Webb, Eugene J., Donald Campbell, Richard Schwartz, Lee Sechrest, and Janet Groves. 1981. *Nonreactive Measures in the Social Sciences*. Boston: Houghton-Mifflin.

*Welch, Eric, Allan Mazur, and Stuart Bretschneider. 2000. "Voluntary Behavior by Electric Utilities: Levels of Adoption and Contribution of the Climate Challenge Program to the Reduction of Carbon Dioxide." *Journal of Policy Analysis and Management* 19 (3): 407.