Strategic Management Journal
Strat. Mgmt. J., **31**: 917–945 (2010)
Published online EarlyView in Wiley InterScience (www.interscience.wiley.com) DOI: 10.1002/smj.840
Received 12 November 2008; Final revision received 19 December 2009

HOW FIRMS RESPOND TO BEING RATED

AARON K. CHATTERJI[1] and MICHAEL W. TOFFEL[2]*
[1] *Fuqua School of Business, Duke University, Durham, North Carolina, U.S.A.*
[2] *Harvard Business School, Boston, Massachusetts, U.S.A.*

While many rating systems seek to help buyers overcome information asymmetries when making purchasing decisions, we investigate how these ratings also influence the companies being rated. We hypothesize that ratings are particularly likely to spur responses from firms that receive poor ratings, and especially those that face lower-cost opportunities to improve or that anticipate greater benefits from doing do. We test our hypotheses in the context of corporate environmental ratings that guide investors to select 'socially responsible,' and avoid 'socially irresponsible,' companies. We examine how several hundred firms responded to corporate environmental ratings issued by a prominent independent social rating agency, and take advantage of an exogenous shock that occurred when the agency expanded the scope of its ratings. Our study is among the first to theorize about the impact of ratings on subsequent performance, and we introduce important contingencies that influence firm response. These theoretical advances inform stakeholder theory, institutional theory, and economic theory. Copyright © 2010 John Wiley & Sons, Ltd.

INTRODUCTION

Information asymmetry has long been understood to complicate market transactions (Akerlof, 1970). Incomplete information prevents buyers from knowing when to believe suppliers' claims about product attributes that are not directly observable prior to purchase. Independent agencies that rate and rank products and companies can help consumers overcome information asymmetries. Such agencies operate in a wide variety of contexts, rating consumer products (Consumer Reports), services (Michelin's guidebooks), and corporate debt (Moody's).[1] These rating schemes are institutions designed to achieve a common objective: to provide credible information to help company stakeholders such as potential buyers, employees, and investors overcome an information disadvantage. Better informed stakeholders can make better decisions about which products to purchase, in which stocks or bonds to invest, and with which companies to seek employment.

Prior scholarship has found evidence that independent company ratings can affect the behavior of consumers and investors. However, scholars have

Keywords: information disclosure; environmental performance; corporate social responsibility; industry self-regulation; ratings
* Correspondence to: Michael W. Toffel, Harvard Business School, Soldiers Field, Boston, MA 02163, U.S.A.
E-mail: mtoffel@hbs.edu

[1] Companies are subjected to an increasing number of ratings and rankings, from 'Best Places to Work' (*Fortune*, 2008; HRC, 2008) to assessments of environmental and social responsibility (Chatterji and Levine, 2006). In fact, a recent survey counted more than 183 public lists across 38 countries of companies rated or ranked on the basis of their reputation for corporate citizenship, employee relations, leadership, innovation, and other characteristics (Fombrun, 2007).

only begun to theorize how independent company ratings affect the organizations being rated, and have offered little guidance on how differences in firm characteristics influence response. For example, with few exceptions (Konar and Cohen, 1997; Lenox and Eesley, 2009), stakeholder theory has emphasized how the identity of stakeholders and the nature of their requests influence firm responsiveness (Eesley and Lenox, 2006; Mitchell, Agle, and Wood, 1997).[2] Institutional theory argues that the legitimacy, and possibly survival, of an organization is threatened when negative information is disclosed about its operations (Hunter and Bansal, 2007). But institutional theory does not offer clear predictions about which firms are most likely to respond to negative information. Finally, economic theory has long appreciated the importance of information disclosure programs in mitigating information asymmetry, but has only recently considered the organizational factors that influence which firms respond (Jin and Leslie, 2009).

We extend existing theories by articulating new explanations for why firms respond to ratings. We argue that because their fortunes rise and fall with key stakeholder groups, companies have an interest in their ratings. Specifically, because company ratings reduce information asymmetry between companies and their stakeholders, managers are motivated to pay attention and, in some contingencies, respond to ratings. In this paper, we theorize and empirically test two important contingencies that condition how firms respond to company ratings.

Specifically, our research examines how firms respond to corporate environmental ratings meant primarily to guide individuals and fund managers who want to invest in environmentally responsible companies. We propose that these ratings, beyond their stated objective of influencing investors, also influence the rated firms. We argue that just as government disclosure regulations that require firms to disclose potentially embarrassing information can lead firms to alter their behavior (Graham, 2000), managers can also be spurred to respond to poor ratings that shame their firms by implementing practices aimed at improving their firms' standing with the independent rating agencies. We further propose that the subset of poorly rated firms that face lower-cost improvement opportunities and greater potential benefits will be especially likely to make the investments needed to improve their ratings. We test our hypotheses using ratings data from one of the foremost rating agencies that discloses assessments of firms' corporate social performance. We examine how hundreds of organizations responded to being involuntarily included when the agency expanded the number of firms it rated.

We make several important theoretical and empirical contributions to the extant literature. Our study is among the first to theorize about how organizations adjust their performance in response to ratings. We hypothesize that two important contingencies, namely, firm-level efficiency and the regulatory environment, moderate how organizations respond, and exploit an exogenous shock to avoid selection issues that can confound empirical evaluations of the effects of ratings. Rarely used in prior research in this domain, this approach constitutes an important empirical contribution to the extant literature. We find evidence that firms initially rated poorly subsequently improved their performance more than two groups of comparison firms: those that were never rated, and those that were initially rated more favorably. We find that this main effect was driven by firms in industries that face significant environmental regulations and by firms that faced less costly opportunities to improve. Our results provide insights for management scholars and policymakers alike who seek to understand how firms respond to public and private regulatory schemes.

LITERATURE REVIEW

Independent company rating and ranking schemes

Prior scholarship on independent company rating and ranking schemes has examined the extent to which they fulfill their primary objective of influencing consumers' and investors' decisions. Some studies have found that investors (Becchetti, Ciciretti, and Hasan, 2007; Rock, 2003) and consumers (Sen and Bhattacharya, 2001) respond to social information disclosure, while other studies yielded mixed or no results (Curran and Moran, 2007; Takeda and Tomozawa, 2008).

[2] Specifically, Konar and Cohen (1997) find that firms whose market value declined most significantly in response to the mandatory disclosure of emissions information were subsequently most likely to reduce their emissions. Eesley and Lenox (2009) find that less polluting firms were more likely to comply with environmental activist requests.

DOI: 10.1002/smj

We are aware of only two studies that have examined how firms respond to independent agents' ratings and rankings, both of which focused on graduate schools' responses to rankings. Elsbach and Kramer (1996) investigated how deans, professors, and students at 'top 20' business schools personally reacted to changes in their schools' *Business Week* rankings. They found that individuals at schools ranked toward the bottom of this elite list felt threatened by the ranking, and deflected the threat by (1) arguing that key dimensions of their schools' strengths were omitted from the ranking criteria, and (2) referring to comparison groups that raised their ranking or status.

Espeland and Sauder (2007) studied law schools' responses to *U.S. News and World Report* rankings. Their study was also based on interviews with deans and faculty members, but the sample included schools from a much broader spectrum. The main effect of a poor ranking was to diminish a school's attractiveness to external funders and high-quality applicants and, in some cases, university presidents responsible for allocating resources. These effects precipitated a 'self-fulfilling prophesy' whereby poor rankings impeded enlistment of the personnel and resources needed to deliver high-quality education. The authors also found evidence that rankings affected management decisions within law schools. In particular, school administrators began to consider how management decisions such as changes in their admissions criteria might affect their rankings.

Although they shed light on how organizations respond to independent rankings and ratings, these studies leave many questions unanswered. It is unclear to what extent these findings, being based on nonprofit organizations' responses to ratings, are generalizable to for-profit companies. Perhaps most important, neither study directly examined how their rankings affected the schools' performance. We address these gaps in the prior literature by examining how responses to a wide range of ratings were reflected in changes in performance in hundreds of companies across a variety of industries.

Government mandatory information disclosure programs

Our research also relates to policy analyses of government mandatory information disclosure programs that require organizations to disclose activities that pose risk. These transparency regulations 'rely on responses to new information by users whose subsequent actions create market or political incentives for disclosers' to modify their behavior (Weil *et al.*, 2006: 158). Many of these transparency regulations thus seek, like independent rating agencies, to directly influence the behavior of organizations' *stakeholders*. Empirical studies have identified several instances in which this has occurred including among investors in, and home owners living near, companies that were required to disclose toxic chemical pollution (Hamilton, 1995; Khanna, Quimio, and Bojilova, 1998; Oberholzer-Gee and Mitsunari, 2006).

Many of these regulations have the additional explicit objective of leveraging stakeholder responses to 'change the practices of targeted organizations in order to achieve specified policy aims' (Weil *et al.*, 2006 :158). For example, restaurant grade cards based on health inspections seek not only to reduce health risks posed to consumers by unhygienic practices, but also to create stronger incentives for restaurant operators to maintain high standards of hygiene (Jin and Leslie, 2003). Similarly, regulations that require factories to report toxic chemical pollution are intended not just to satisfy communities' 'right to know' about the toxins in their environment but also to pressure factories to reduce their emissions.

A number of studies have found that organizations do respond to government mandatory information disclosure programs. After finding that a state regulation requiring companies to warn consumers about toxic materials in their products inspired what Graham (2000) called 'a flurry of efforts' to reduce or eliminate these materials, Graham concluded that 'regulation by shaming' was 'a newly potent political force' (Graham, 2000: 36). Similarly, Bennear and Olmstead (2008) found that a regulation that mandates disclosure to customers of information about regulatory violations and contaminant levels led many utilities to improve their regulatory compliance. Other studies have found government information disclosure programs have spurred companies to improve environmental performance (Blackman, Afsah, and Ratunanda, 2004; Konar and Cohen, 1997; Scorse, 2007), food and water safety (Bennear and Olmstead, 2008; Jin and Leslie, 2003), and surgical outcomes (Cutler, Huckman, and Landrum, 2004; Hannan *et al.*, 1994; Peterson *et al.*, 1998).

DOI: 10.1002/smj

These empirical results suggest that information about a company's management practices and performance disclosed pursuant to government regulations can stimulate management to pursue changes in both. Our paper is among the first to explore whether such information, when disclosed by non-governmental, independent rating agencies, might similarly motivate management to improve practices and performance. Furthermore, we look beyond average effects to identify several contingencies where such effects are likely to be more pronounced.

FIRMS' RESPONSES TO INDEPENDENT RATINGS

Responding to poor ratings

By defining widely accepted standards of behavior and comparing organizations' adherence to these standards, independent rating agencies help prospective consumers, employees, and investors identify which organizations possess high-quality but difficult-to-observe management practices. The financial credit rating firm Moody's, for example, describes itself as providing 'credit ratings and research [to] help investors analyze the credit risks associated with fixed-income securities' (Moodys.com, 2008). Similarly, *U.S. News & World Report* maintains that its college and graduate program rankings are intended not to 'transform law schools' or 'hold them accountable,' but rather to 'provide accessible information to educational consumers' (Espeland and Sauder, 2007: 5).

In the face of growing investor interest in 'socially responsible investing,' and a desire by some to avoid investing in firms deemed socially irresponsible (Barnett and Salomon, 2006: 1101), social rating agencies emerged to 'identify which firms are more or less responsible' (Vogel, 2005: 39). Such agencies conduct in-depth analyses of companies' management practices and social performance by interviewing company managers, reviewing corporate information and media reports, and synthesizing company records from various regulatory agencies. The social rating firm KLD Research & Analytics, the focus of our empirical analysis, describes its mission purely in terms of serving investors, specifically, as 'providing management tools to professionals integrating environmental, social and governance factors (ESG) into their investment decisions' (KLD.com, 2006).

Granting that pressuring companies is not the explicit mission of social rating agencies, we argue that their ratings nevertheless elicit responses from some companies. Like Rao (1994: 32), we view these ratings as 'social tests,' where favorable ratings bestow a high status upon firms, which are then presumed to be superior to other firms on the dimension of interest. In contrast, poorly rated firms' management practices and performance are construed to lie outside the boundaries deemed by the rating agency to be 'desirable, proper, [and] appropriate' (Bansal and Clelland, 2004: 94).

Poor ratings can prompt managerial action to address concerns evoked internally, within their organizations, as well as concerns that develop among external stakeholders. A poor reputation for environmental performance and corporate social responsibility can undermine employee morale (Ramus and Killmer, 2007; Savitz and Weber, 2007), innovativeness, and willingness to engage in participatory problem solving (Ramus and Steger, 2000). Managers of poorly rated firms are also more likely to suffer 'public humiliation' (Graham, 2000: 36) and become motivated to improve their companies' ratings by implementing more robust environmental management practices. Stephan (2002: 194) argued that a facility's pollution levels 'may signal to the market the overall economic health of a given industrial plant. . .[and] inefficient pollution output may signal reduced profit margins, increased liabilities, and ineffective management.' Similarly, a poor environmental rating can trigger concerns by current and potential investors that the firm might be overlooking opportunities to invest in environmental management activities that might reduce costs, preempt competition, and spur higher-order learning (Hart, 1995; King and Lenox, 2002; Sharma and Vredenburg, 1998). Firms with poor environmental ratings might also be perceived to be operating with unusually high risks of accidents and resulting liability exposure (Delmas, 2002). Investors might also perceive firms with poor environmental ratings to face higher risks of business interruptions and legal costs, inasmuch as poor ratings can erode relationships with regulators and local communities (Coglianese and Nash, 2001; Delmas and Toffel, 2008).

More broadly, poor environmental ratings might sully companies' overall reputations. Hamilton (1995) found that when new environmental performance information was disclosed, the firms with

DOI: 10.1002/smj

the worst records were more likely to attract negative media coverage. Beyond reputational concerns, negative media coverage of environmental issues is associated with greater stock price volatility (Bansal and Clelland, 2004). Poor environmental reputations are also likely to put companies on environmental activists' radar screens and make them targets of lawsuits, protests, boycotts, letter writing campaigns, and proxy votes (Lenox and Eesley, 2009). Companies with poor environmental ratings also risk alienating buyers who incorporate environmental and social considerations into their procurement criteria, and risk being excluded from the rapidly growing market for *green funds* (Norton, 2007). All of these factors can lead companies to fear that poor environmental ratings might erode their stock market value, which was the case for some companies revealed by a government information disclosure program to have poor environmental performance (Hamilton, 1995; Khanna *et al.*, 1998). Moreover, empirical evidence has revealed that firms with poor KLD social and environmental ratings suffered below-average market returns (Kempf and Osthoff, 2007).

The growing interest in corporate social responsibility and socially responsible investment has increased both the salience of independent ratings agencies and companies' responsiveness to risks to their brand reputations. The *Financial Times* noted that 'in post-industrial society, brands have replaced factories as companies' most valuable assets.... [C]hief executives...dare not risk damaging their brands by being seen as hostile to people or the planet' (Tomkins, 2001). According to Gunningham, Kagan, and Thornton (2004: 308), 'corporate executives increasingly talk about the importance of...[avoiding] activities that societies (or influential elements within them) deem unacceptable.' In his comprehensive assessment of the literature on corporate social responsibility, Vogel (2005: 52) observed that '[m]any companies now regard it as in their self-interest to be, or at least appear to be, responsive to [non-governmental organization] and media criticism, lest their reputations suffer significant damage.' Even the mere threat of protest campaigns by non-governmental organizations has prompted many companies to make policy changes and enact more stringent social and environmental management practices (Vogel, 2005).

The above arguments imply that firms that receive poor ratings are more likely to respond by taking management actions to bolster their ratings. Because environmental ratings are based largely on environmental performance, such management actions will focus on improving environmental performance. Taken together, this implies that the worse a firm's initial environmental rating, the more likely its management will respond with actions to improve the firm's environmental performance. As a result, we predict that:

> *Hypothesis 1: Firms that receive a poor environmental rating will subsequently improve their environmental performance more than other firms will.*

Differential benefits of responding to poor ratings

We do not, however, expect firms to respond uniformly to poor environmental ratings. The benefits from and costs of responding to negative ratings vary, in part, due to differences across the regulatory environments firms face. Moreover, these environments are dynamic, with regulatory thresholds and enforcement stringency changing over time, and the cost of complying with government regulations varying widely across industries (Leone, 1981, 1986). Firms in industries subject to significant environmental regulations, for example, 'face greater exposure to the public policy process' (Cho and Patten, 2007: 642), and face a higher risk that a poor rating might provoke negative media coverage and concern within the communities surrounding their plants. Both of these forces can stimulate political pressure to increase the frequency and/or intensity (and thus cost) of regulatory inspections. Firms that face greater regulatory threats are thus more likely to invest in environmental improvements (Cho and Patten, 2007; Lyon and Maxwell, 2001; Short and Toffel, 2008), establish industry self-regulation schemes, and adopt voluntary environmental programs (Corbett, Montes-Sancho, and Kirsch, 2005; King and Lenox, 2000). We propose that firms that face significant environmental regulations will be especially motivated by poor environmental ratings to improve their environmental performance.

> *Hypothesis 2: Within industries that face significant environmental regulations, firms that receive a poor environmental rating will subsequently improve their environmental performance more than other firms will.*

DOI: 10.1002/smj

Differential costs of responding to poor ratings

We also expect that some firms will be able to make environmental improvements at lower cost. Just as firms face different costs and benefits of complying with environmental regulations (Terlaak, 2007), they also face different sets of opportunities for improving their environmental performance (Levi and Nault, 2004) and possess different capabilities that affect the cost of pursuing these opportunities. Firms that have already made substantial investments in mitigating their environmental impact often find that additional improvements require the adoption of increasingly costly technologies and management programs (Darnall and Edwards, 2006; Graham and Miller, 2001; Hart and Ahuja, 1996).

Hart and Ahuja (1996: 32) argue that cheap 'changes that result in large emission reductions relative to costs' are more likely to be available to the least environmentally efficient firms: those with poor environmental performance given their size. Such firms typically have a greater opportunity to exploit *low-hanging fruit* (Darnall and Edwards, 2006; King and Lenox, 2000; Reinhardt, 1998; Terlaak, 2007). According to Terlaak (2007: 977) '[F]irms with substandard practices have more opportunities to exploit low-hanging fruit' because they face lower marginal costs of improving their performance.

Laggards might further benefit by leveraging the experience of leaders. Technological laggards, von Hippel (1988) observes, can learn from leaders in the field, borrow off-the-shelf technologies, or tap existing internal know-how at far lower cost than firms that have already achieved superior performance. We believe the same mechanisms apply to environmental technologies and management techniques that can improve environmental performance.

We thus extend our first hypothesis, that poorly rated firms will subsequently improve their environmental performance more than will other firms, by suggesting that this relationship is especially likely to be present among less environmentally efficient firms because they are more likely to be able to exploit lower cost opportunities to improve.

Hypothesis 3: Among less environmentally efficient firms, those that receive a poor environmental rating will subsequently improve their environmental performance more than other firms will.

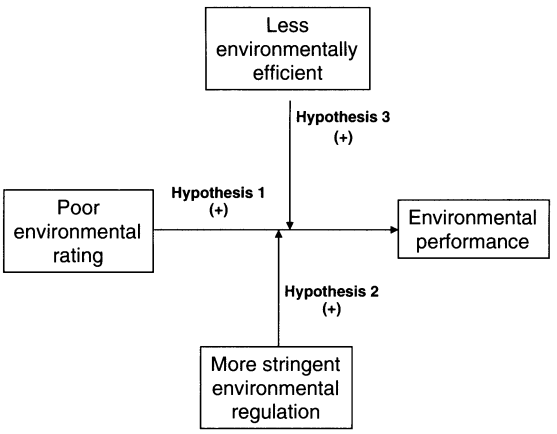

Figure 1. Conceptual model

Figure 1 graphically depicts our three hypotheses.

DATA AND MEASURES

Environmental performance

We measure environmental performance using corporate-wide toxic pollution, an outcome metric employed by many other scholars (Delmas, Russo, and Montes-Sancho, 2007; Kassinis and Vafeas, 2006; King and Lenox, 2002; Klassen and Whybark, 1999; Russo and Harrison, 2005). Specifically, we use the total pounds of toxic chemical emissions each firm reported to the U.S. Environmental Protection Agency's (EPA) Toxic Release Inventory (TRI) as production waste, transfers, and releases. *Toxic emissions* based on TRI data are among the most commonly used outcome measures of environmental performance, in part because the data are legally required to be disclosed in a consistent manner across a wide array of industries.[3] We obtained TRI data from the Corporate Environmental Profiles Directory (CEPD) created by the Investor Responsibility Research Center, which aggregates facility-level data from a variety of EPA databases for all domestic subsidiaries of all members of the S&P (Standard & Poor's) 500 Index, S&P SmallCap 600 Index, and

[3] Whereas some studies apply various weights to these chemicals to account for differences in toxicity, simply summing the pounds of emissions was a method commonly used by the media and prominent nonprofit organizations and in government publications during the sample period (Toffel and Marshall, 2004).

DOI: 10.1002/smj

S&P MidCap 400 Index. To reduce the impact of outliers on our results, we take the log after adding 1, a common practice in empirical analyses that employ TRI data (Kassinis and Vafeas, 2006; King and Lenox, 2000; Russo and Harrison, 2005).

Environmental ratings

We obtained environmental ratings from KLD Research & Analytics, Inc. (KLD), 'the largest multidimensional CSP [corporate social performance] database available to the public' (Deckop, Merriman, and Gupta, 2006: 334). KLD, which has been issuing environmental ratings for all members of the S&P 500 Index and Domini Social 400 Index since 1991, collects and analyzes data from five major sources: direct communication with company managers, KLD's research partners around the world, the media, public documents, and government and non-governmental organizations. Its ratings are thus based on publicly available information as well as information collected directly from the rated companies. As part of its ratings process, KLD sends its ratings to senior managers of the rated firms to prompt discussions about potential discrepancies.

KLD expanded its coverage in 2001, when it began including ratings for Russell 1000 Index members in its KLD STATS database. Ratings for Russell 2000 Index members were added to KLD STATS in 2003. Because KLD's decision to begin rating these firms was unrelated to their behavior or performance, and because the firms had no influence on the decision to be rated, we avoid the selection problems common to many program evaluations.

KLD ratings are widely known in social investing circles: 15 of the world's top 25 institutional financial managers use KLD research, and more than $10 billion is invested in funds based on KLD's ratings (KLD.com, 2006). For example, the Teachers Insurance and Annuity Association-College Retirement Equities Fund (TIAA-CREF) uses KLD ratings as the basis for including equities in its Social Choice Equity fund (Baue, 2003), which had nearly $700 million in assets under management in 2009 (TIAA-CREF, 2008, 2009). When KLD downgraded its rating of The Coca-Cola Company in 2006 due to concerns about its labor and environmental practices in the developing world, TIAA-CREF divested more than $50 million worth of the company's stock (Wilbert, 2006). Widely used in studies of corporate social responsibility and socially responsible investing (Berman *et al.*, 1999; Margolis and Walsh, 2003), KLD ratings have been referred to as 'the *de facto* research standard' in those domains (Waddock, 2003: 369). Moreover, prior research has found empirical support for the construct validity (Sharfman, 1996) and predictive validity (Chatterji, Levine, and Toffel, 2009) of KLD's ratings.

We obtained annual company environmental ratings data for each of KLD's 14 dichotomous environmental 'strength' and 'concern' variables from the KLD STATS database. The seven environmental 'strength' variables are: beneficial products and services; pollution prevention; recycling; clean energy; communications; property, plant, and equipment; and other strengths. The seven environmental 'concern' variables are: hazardous waste; regulatory problems; ozone-depleting chemicals; substantial emissions; agricultural chemicals; climate change; and other concerns. Detailed descriptions of these ratings are provided in Exhibit A2 of the Appendix.

Following Cho and Patten (2007), we created a dummy variable, *initial rating poor*, to indicate firms that had initial KLD environmental ratings consisting only of concerns (no strengths). Such a rating suggests corporate behavior that violates taken-for-granted norms (Scott, 1987). Another dummy variable, *initial rating mixed or good*, was created to identify firms that had initial environmental ratings consisting only of strengths (no concerns), of both strengths and concerns, or of neither strengths nor concerns. In our empirical analysis, we interact these firm-level variables with a time-varying dummy variable, *KLD rated*, coded '1' in years in which firms were rated by KLD (regardless of the actual rating), and '0' otherwise.

Regulatory scrutiny

We identified firms that operate in a highly environmentally regulated context based on a classification devised by Cho and Patten (2007). A dichotomous variable, *highly environmentally regulated*, was coded '1' for companies with a primary standard industrial classification (SIC) code of mining (SIC 10), oil exploration (13), paper (26), chemical and allied products (28), petroleum refining (29), metals (33), or utilities (49), and '0' otherwise. We created a second dichotomous variable, *low environmentally regulated*, to which we

DOI: 10.1002/smj

applied the inverse of the aforementioned coding scheme.

Environmental efficiency

Environmental efficiency (or 'eco-efficiency') refers to the extent to which a firm's environmental impacts or resource intensity are minimized, normalized by its production level (Ayres, 1995; Starik and Marcus, 2000). We operationalize this as the ratio of each firm's toxic chemical emissions to revenues. We obtained these data from CEPD and Compustat, respectively. We calculated each firm's average ratio during 1999–2000, the two-year period immediately before any of the firms in our sample were rated by KLD, and compared these ratios to the corresponding industry median value during that period. This yielded two firm-level dummy variables: *less environmentally efficient* was coded '1' for firms with a ratio that exceeded the industry median, and '0' otherwise; *more environmentally efficient* was coded '1' for firms with ratios less than the industry median, and '0' otherwise.

Control variables

We control for several other factors that might influence environmental performance including regulatory context (Delmas and Toffel, 2008) and organization size (Goodstein, 1994). We control for regulatory context by including firm-level fixed effects and year dummies, and for organization size by including the logarithmic transformations of annual *employment, revenues*, and *assets* (Christmann, 2000; King, Lenox, and Terlaak, 2005; Russo and Fouts, 1997; Sharma, 2000; Waddock and Graves, 1997), obtained from Compustat. Because a firm's acquisitions or divestitures of TRI-reporting facilities can also affect its aggregate TRI emissions, we control for the *number of TRI-reporting facilities,* obtained from the CEPD.

METHODS AND RESULTS

Because we are interested in how firms respond to their initial KLD ratings, we compare firms first rated consequent to KLD's expansion with firms not rated by KLD during our sample period. Including the latter enables us to control for performance changes attributable to the availability of new technologies or changes in regulations that could affect the environmental performance of all, not just the rated, firms.

Our analysis begins in 1999, two years before KLD expanded its scope, and extends through 2004 (the most recent data available from CEPD). Our sample includes 598 companies representing a wide variety of industries (Table 1). Sample firms meet *all* of the following criteria: owned at least one EPA-regulated facility; were members of the S&P SmallCap 600 or S&P MidCap 400 Index; and were not listed in the Domini Social 400 Index prior to 2001. Summary statistics and correlations are provided in Table 2.

Descriptive results

We first examine the raw data to look for evidence of whether firms' emissions levels change after being rated. Normalizing corporate-wide toxic chemical emissions (in pounds) by accounting for corporate sales (in dollars) and the number of TRI-reporting facilities, and taking the log to reduce the highly skewed distribution, we find that for firms initially rated poor this ratio averaged 6.2 in the pre-rating period and 4.9 in the post-rating period. This 26 percent decline was 70 percent larger than that experienced by firms initially rated good or mixed, for which this ratio decreased from 4.6 in the pre-rating period to 4.01 in the post-rating period. This is a statistically significant as well as a substantive difference in trends.[4]

Empirical models

We test our hypotheses using a difference-in-differences approach to compare firms' environmental performance before and after being rated, and use as a reference group firms that were not rated.

[4] To determine whether the difference in trends was statistically significant, we used OLS pooled regression (with robust standard errors clustered by firm) to estimate the following model: $y = \beta_1\ p_i + \beta_2\ mg_i + \beta_3\ p_i \times R_{it} + \beta_4\ mg_i \times R_{it}$ where y = log (emissions/[sales × facilities]), p_i is a firm-level dummy coded '1' for firms initially rated poor, mg_i = is a firm-level dummy coded '1' for firms initially rated mixed or good, and R_{it} is a dummy coded '1' for years in which a firm was rated by KLD. The statistical significance of the difference in trends was determined using a Wald test of the equality of $\beta_3 = \beta_4$, which yielded F=3.66, p=0.06.

DOI: 10.1002/smj

Table 1. Sample statistics

Panel A: Number of firms in sample

	1 Total[a]	2 Less environmentally efficient	3 More environmentally efficient
Firms never rated	240	59	79
Firms initially rated mixed or good	297	127	105
Firms initially rated poor	61	43	14
Total number of firms	598	229	198

[a] The sample of firms used to test Hypothesis 1 (results in Table 2) is depicted in Column 1. The sample of firms used to test Hypothesis 3 (results in Table 4) is depicted in Columns 2 and 3. The former exceeds the latter because classifying firms as more or less environmentally efficient is based on emissions and revenue data from 1999–2000, which not all firms in Column 1 reported.

Panel B: Industry composition of sample

NAICS code (three-digit)	Description	Number of firms
334	Computer and electronic product manufacturing	98
325	Chemical manufacturing	60
336	Transportation equipment manufacturing	49
333	Machinery manufacturing	43
331	Primary metal manufacturing	32
221	Utilities	31
332	Fabricated metal product manufacturing	29
339	Miscellaneous manufacturing	25
311	Food manufacturing	24
335	Electrical equipment, appliance, and component manufacturing	21
322	Paper manufacturing	19
212	Mining (except oil and gas)	13
324	Petroleum and coal products manufacturing	12
327	Nonmetallic mineral product manufacturing	12
326	Plastics and rubber products manufacturing	11
541	Professional, scientific, and technical services	11
423	Merchant wholesalers, durable goods	10
Various	Other industries	98
	Total	598

We test Hypothesis 1 by estimating the following model:[5]

$$Y_{i,t} = \beta_1\ KLD\ rated_{i,t} \times initial\ rating\ poor_i + \beta_2\ KLD\ rated_{i,t} \times initial\ rating\ mixed\ or\ good_i + \beta_3\ X_{i,t} + \beta_4 \gamma_t + \alpha_i + e_{i,t} \quad (1)$$

$Y_{i,t}$ refers to the TRI emissions of firm i in year t, and $X_{i,t}$ includes the log of assets, revenues, employment, and number of TRI-reporting facilities. Firm-level fixed effects (α_i) control for time-invariant factors during our sample period (such as a firm's corporate culture and geographic location), and a full set of year dummies (γ_t) accounts

[5]We include in our specifications two interaction terms to facilitate interpretation of the regression coefficients. The coefficient on the first interaction term directly tests Hypothesis 1. Our specification is interchangeable with a specification that instead

DOI: 10.1002/smj

Table 2. Descriptive statistics

Panel A: Summary statistics

Variable	Mean	SD	Min	Max
1. Log pounds of emissions	12.23	4.18	0	20.71
2. KLD rated × initial rating poor	0.06	0.24	0	1
3 KLD rated × initial rating mixed or good	0.26	0.44	0	1
4. Log employees	8.63	1.46	1.95	13.09
5. Log sales	20.97	1.55	15.13	26.38
6. Log assets	21.04	1.63	16.93	27.74
7. Log number of TRI-reporting facilities	1.65	0.87	0.69	4.76

Note: 2,412 firm-year observations.

Panel B: Correlations

	1	2	3	4	5	6
1. Log pounds of emissions	1.00					
2. KLD rated × initial rating poor	0.09	1.00				
3 KLD rated × initial rating mixed or good	−0.09	−0.15	1.00			
4. Log employees	0.24	−0.01	−0.06	1.00		
5. Log sales	0.31	0.07	−0.07	0.89	1.00	
6. Log assets	0.30	0.09	−0.06	0.83	0.94	1.00
7. Log number of TRI-reporting facilities	0.52	0.07	−0.04	0.45	0.46	0.40

Note: 2,412 firm-year observations.

for annual technological and policy changes that might affect emissions. Hypothesis 1 predicts that β_1 will be negative, indicating that environmental performance improved more at firms for which the initial KLD rating was poor than at unrated firms. It also predicts that β_1 will be significantly smaller than β_2, indicating that environmental performance improved more at firms for which the initial KLD rating was poor than at firms for which the initial KLD rating was mixed or good.

To test the moderating effects described in Hypotheses 2 and 3, we estimate models similar to Equation 1, but fully interact all variables with two dummy variables. Our model that tests Hypothesis 2 interacts all variables with *highly environmentally regulated* and *low environmentally regulated*; the model that tests Hypothesis 3 interacts all variables with *less environmentally efficient* and *more environmentally efficient*. These specifications, like the specification that tests Hypothesis 1, identify changes in performance levels between firms initially rated poor, firms initially rated mixed or good, and unrated firms. Our interaction terms enable us to make comparisons *within* these additional subcategories.[6] For example, in testing Hypothesis 3, our estimates compare performance *within* the less environmentally efficient subset of firms, and *within* the more environmentally efficient subset.

Empirical results

We estimated our models using Stata, employing ordinary least squares (OLS) regression with firm-level fixed effects. Our estimation technique is predicated on the assumption that the environmental performance of each group of newly rated firms would have followed the trend of the unrated firms had KLD not expanded the scope of its coverage. Although not directly testable, this assumption would be strengthened if the performance trends of these three groups were found to be similar during the pre-rating period. To test this, we compared trends from 1999 to 2000, the period before

includes the main effect (*KLD rated*) and one of these interactions (e.g., *KLD rated* × *initial rating poor*), except with that specification the two OLS coefficients must be added together to determine whether the effect of poor ratings differed from the performance of the unrated firms (as predicted by Hypothesis 1).

[6] These empirical specifications are virtually identical to running separate regressions on split samples (e.g., the less environmentally efficient subsample and then the more environmentally efficient subsample), but employing a single, fully interacted model facilitates comparing coefficients.

Table 3. Performance improved most among firms initially rated poor
Dependent variable: log toxic emissions

		1	2
(A)	KLD rated × initial rating poor	−1.305***	−0.656***
		[0.229]	[0.254]
(B)	KLD rated × initial rating mixed or good	−0.314**	0.315*
		[0.122]	[0.170]
	Log employees	0.181	−0.313
		[0.240]	[0.241]
	Log sales	0.310	0.718***
		[0.234]	[0.234]
	Log assets	−0.580**	−0.181
		[0.267]	[0.265]
	Log number of TRI-reporting facilities	1.824***	1.778***
		[0.144]	[0.142]
	Year 2000		−0.131
			[0.142]
	Year 2001		−0.476***
			[0.145]
	Year 2002		−0.520***
			[0.159]
	Year 2003		−0.534***
			[0.201]
	Year 2004		−1.666***
			[0.208]
	Firm fixed effects	Included	Included
	Observations (firm-years)	2412	2412
	Firms	598	598
	R-squared (within)	0.12	0.16
	Wald test: coefficient on (A) = (B)?	15.58***	15.63***

OLS regression coefficients, with standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.10. Wald test displays F test statistic, where null hypothesis is that the coefficients are statistically indistinguishable. The sample includes newly rated firms and never-rated firms. Column 1 displays results of a model that omits year dummies, where the interaction terms estimate the *absolute change* in emissions from the pre- to the post-rating periods. Column 2 includes year dummies to account for temporal shocks that affect the entire sample, including the control (never-rated) firms. Here, the interaction terms estimate the *relative change* in emissions, from the pre- to the post-rating periods, between the rated group and the control (never-rated) firms.

any of the firms in our sample were rated. T-tests performed to compare the percent change in emissions per sales from 1999 to 2000 revealed the pre-period trends of the three focal groups—those eventually rated poor, those eventually rated mixed or good, and those never rated—to be statistically indistinguishable.

Table 3 presents the results of the model that tests Hypothesis 1. For context, we first estimate a simpler specification that omits year dummies in which the coefficients on the two interaction terms estimate the *absolute* change in emissions between the pre- and post-rating periods. Note that the coefficients on both the *KLD rated × initial rating poor* and the *KLD rated × initial rating mixed or good* variables are negative and statistically significant, indicating that for both groups emissions declined after the firms were rated (Column 1).[7] But because these coefficients were estimated in a model that omitted year dummies, these absolute differences fail to account for emission declines that were also experienced by the control group of firms that were never rated. Our main model includes year dummies to account for the annual shocks that affect all firms in the sample, and yield the difference-in-differences estimates that test Hypothesis 1 (Column 2). The results of our main model indicate that firms initially rated poor subsequently reduced their emissions by 0.66 log points more than firms never rated ($\beta = -0.66$; p=0.01), a magnitude equal to

[7] A Wald test indicates that emissions declined significantly more for firms rated poor than for firms rated mixed or good (F=15.58, p<0.01).

Table 4. Rating effects moderated by regulatory stringency
Dependent variable: log toxic emissions

		1	2
(A)	Highly environmentally regulated × KLD rated × initial rating poor	−1.538***	−1.027***
		[0.294]	[0.348]
(B)	Highly environmentally regulated × KLD rated × initial rating mixed or good	0.178	0.717**
		[0.240]	[0.313]
	Highly environmentally regulated × log employees	−0.06	−0.962*
		[0.516]	[0.548]
	Highly environmentally regulated × log sales	0.316	0.750**
		[0.360]	[0.369]
	Highly environmentally regulated × log assets	−0.318	0.328
		[0.511]	[0.524]
	Highly environmentally regulated × log number of TRI-reporting facilities	1.761***	1.725***
		[0.257]	[0.257]
(C)	Low environmentally regulated × KLD rated × initial rating poor	−0.975***	−0.211
		[0.376]	[0.397]
(D)	Low environmentally regulated × KLD rated × initial rating mixed or good	−0.474***	0.238
		[0.142]	[0.205]
	Low environmentally regulated × log employees	0.243	−0.149
		[0.273]	[0.272]
	Low environmentally regulated × log sales	0.261	0.651**
		[0.308]	[0.308]
	Low environmentally regulated × log assets	−0.611*	−0.241
		[0.321]	[0.319]
	Low environmentally regulated × log number of TRI-reporting facilities	1.866***	1.826***
		[0.174]	[0.172]
	Year dummies (2001–2004)		Included
	Firm fixed effects	Included	Included
	Observations (firm-years)	2412	2412
	Firms	598	598
	R-squared (within)	0.12	0.17
	Wald test: coefficient on (A) = (B)?	22.62*	24.43***
	Wald test: coefficient on (C) = (D)?	1.61	1.36
	Wald test: coefficient on (A) = (C)?	1.39	2.39
	Wald test: coefficients on (A) − (B) = (C) − (D)?	5.17**	6.13**

OLS regression coefficients, with standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.10. Wald test displays F test statistic, where null hypothesis is that the coefficients are statistically indistinguishable. The sample includes newly rated firms and never-rated firms. See the footnote to Table 3 for the differences between these two models, including differences in how to interpret coefficients on the interaction terms.

one-sixth of one standard deviation (calculated as $\beta = -0.66$ divided by SD of log toxic emissions $= 4.18$). A Wald test that compared the coefficients on the two interaction terms revealed that firms initially rated poor also reduced their emissions more than firms initially rated mixed or good ($F=15.63$; $p<0.01$). These results support Hypothesis 1.[8]

Table 4 presents the results of the fully interacted model that tests Hypothesis 2. Again, for context, Column 1 reports results of the simpler model that omits year dummies. These results indicate that three of the four rated groups experienced *absolute* declines in emissions from the pre-rated period to the period during which they were rated (the fourth group exhibited no significant change). Our main model (Column 2), which includes year dummies, yields results that also account for the control group's temporal trends. Here, the negative statistically significant coefficient on the first interaction term (*Highly environmentally regulated* × *KLD rated* × *initial rating poor*) indicates that

[8] Reestimating our main model using heteroskedasticity robust standard errors or robust standard errors clustered by firm, yielded results similar to our main results: the coefficient on our hypothesized variable, *KLD rated* × *initial rating poor*, remained significantly different from zero and from *KLD rated* × *initial rating good or mixed*.

DOI: 10.1002/smj

highly regulated firms that were initially rated poor reduced emissions by 1.03 log points ($p<0.01$), just under one quarter of one standard deviation (calculated as $\beta = -1.03/SD=4.18$), more than did firms in the control group (of highly regulated firms that were never rated). A Wald test comparing the coefficient on this first interaction term to the coefficient on the second interaction term (*Highly environmentally regulated* × *KLD rated* × *initial rating mixed or good*) revealed that highly regulated firms initially rated poor also reduced emissions to a greater extent than did highly regulated firms initially rated mixed or good (Wald test $F = 24.43$; $p<0.01$). The insignificant coefficients on the third and fourth interaction terms indicate a lack of evidence that either newly rated group of less regulated firms performed any differently than the less regulated firms that were never rated. Taken as a whole, the results presented in Table 4 support Hypothesis 2 by indicating that among highly regulated firms a poor initial rating was *particularly* associated with performance improvement.

As an extension, we considered the relative improvement within highly regulated industries between firms initially rated poor and the remaining firms. We compared that to the relative improvement between these groups of firms within low regulated industries. We found that the difference within the former is greater than the latter, and a Wald test indicated that this difference in differences was statistically significant.[9]

The fully interacted model that tests Hypothesis 3 is estimated on a slightly smaller sample than our earlier models (2,068 versus 2,412 firm-year observations) because we now omit those firms that lacked the 1999 or 2000 emissions and revenue data needed to classify them as more or less environmentally efficient. The simpler model that omits year dummies indicates that all four rated groups exhibited absolute declines in emissions from the pre- to the post-rated period (Column 1 of Table 5). The full model (Column 2) indicates that *among* the *less environmentally efficient* firms, those initially rated poor reduced emissions by 0.77 log points more than firms of this type that were never rated, a difference of just over one-sixth of one standard deviation (calculated as $\beta = -0.77/SD=4.18$). A Wald test revealed that *among* the *less environmentally efficient firms*, those initially rated poor also reduced emissions to a greater extent than firms initially rated mixed or good (Wald test $F = 16.60$; $p<0.01$).[10] For completeness, we also note that the coefficients on the two *KLD rated* variables interacted with the *more environmentally efficient* group in Column 2 provide no evidence that either of the newly rated groups of more efficient firms subsequently performed better than the more efficient firms that were never rated. The results in Table 5 support Hypothesis 3, as they indicate that poor ratings are associated with performance improvement particularly among less environmentally efficient firms.

As an extension similar to the one we conducted earlier, we considered the relative improvement within the set of less environmentally efficient firms between firms initially rated poor and the remaining firms. We compared that to the relative improvement between these groups of firms within the set of more environmentally efficient firms. While we found that the difference within the former is greater than the latter, a Wald test indicated that this difference in differences was not statistically significant.

Robustness tests

We ran a series of tests to assess the robustness of our results. We began by running a falsification test, whereby we reestimated our main difference-in-differences model assigning each firm a false year in which its rating period begins. We focused this exercise on the years 1994–1998, the period

[9] For example, the results of Model 1 indicate that, within the highly regulated industries, the firms initially rated poor reduced their emissions by 1.538 log points whereas the remaining firms increased emissions by 0.178 log points, a difference of 1.706 log points. Within the low regulated industries, the difference in improvements between these groups is 0.501 log points (0.975 versus 0.474). A Wald test, reported in the last row of Table 4, confirmed that the difference between these differences (1.205 log points, 1.706 versus 0.0501) was statistically significant (Wald $F=5.17$, $p<0.05$). Similarly, the difference between these groups was statistically significant in Model 2 (Wald $F=6.13$, $p<0.05$).

[10] Specifically, the Wald test examined whether the coefficient on *Less environmentally efficient* × *KLD rated* × *initial rating poor* statistically differed from the coefficient on *Less environmentally efficient* × *KLD rated* × *initial rating mixed or good.* To overcome concerns that our results might be affected by sales entering our main model both as a control variable and as part of our approach to split the sample into more and less environmentally efficient firms, we estimated as a robustness test an alternative model that omitted the two sales control variables. This alternative model yielded results that were nearly identical to our main results.

Table 5. Rating effects moderated by environmental efficiency
Dependent variable: log toxic emissions

		1	2
(A)	Less environmentally efficient × KLD rated × initial rating poor	−1.461***	−0.772***
		[0.252]	[0.296]
(B)	Less environmentally efficient × KLD rated × initial rating mixed or good	−0.365**	0.360
		[0.161]	[0.233]
	Less environmentally efficient × log employees	0.601*	0.087
		[0.312]	[0.312]
	Less environmentally efficient × log sales	0.113	0.475
		[0.305]	[0.303]
	Less environmentally efficient × log assets	−0.959***	−0.623*
		[0.342]	[0.339]
	Less environmentally efficient × log number of TRI-reporting facilities	1.427***	1.208***
		[0.205]	[0.205]
(C)	More environmentally efficient × KLD rated × initial rating poor	−0.848*	0.100
		[0.442]	[0.460]
(D)	More environmentally efficient × KLD rated × initial rating mixed or good	−0.365**	0.541**
		[0.182]	[0.247]
	More environmentally efficient × log employees	0.109	−0.549
		[0.375]	[0.376]
	More environmentally efficient × log sales	0.500	0.980***
		[0.346]	[0.344]
	More environmentally efficient × log assets	−0.032	0.771*
		[0.423]	[0.425]
	More environmentally efficient × log number of TRI-reporting facilities	1.872***	1.836***
		[0.204]	[0.199]
	Year dummies (2000–2004)		Included
	Firm fixed effects	Included	Included
	Observations (firm-years)	2068	2068
	Firms	427	427
	R-squared (within)	0.13	0.18
	Wald test: coefficient on (A) = (B)?	14.63***	16.60***
	Wald test: coefficient on (C) = (D)?	1.06	0.94
	Wald test: coefficient on (A) = (C)?	1.45	2.54†
	Wald test: coefficients on (A) − (B) = (C) − (D)?	1.24	1.68

OLS regression coefficients, with standard errors in brackets. *** $p<0.01$, ** $p<0.05$, * $p<0.10$, † $p=0.11$. Wald test displays F test statistic, where null hypothesis is that the coefficients are statistically indistinguishable. The sample includes newly rated firms and never-rated firms. See the footnote to Table 3 for the differences between these two models, including differences in how to interpret coefficients on the interaction terms.

before KLD expanded its coverage to the firms in our sample. We assigned a placebo 'rated period' of 1996–1998 to firms that were subsequently rated by KLD during our real sample period. We first estimated placebo regressions without year dummies to assess absolute changes in emissions between the pre-placebo-rated (1994–1995) and post-placebo-rated (1996–1998) periods. The results indicated no change in absolute emissions among firms that ultimately received poor ratings, and a significant decline among firms ultimately rated mixed or good. The results of the full placebo model, which includes year dummies, revealed that over this time period the firms that ultimately received poor KLD ratings decreased emissions less rapidly than both the control (never-rated) firms and firms that ultimately received mixed or good KLD ratings. Viewing these placebo results alongside our main results, we conclude that firms ultimately rated poor by KLD exhibited an inferior performance trend during the mid-1990s, which reversed when they were rated in the following decade.

We also performed robustness tests to assess the extent to which our results were sensitive to plausible alternative measures of our independent and dependent variables, and to changes in our sample. We first investigated whether our results were

DOI: 10.1002/smj

sensitive to the manner in which we categorized firms as being more or less environmentally efficient during the pre-rating period. In our main analysis, we made this categorization based on whether each firm's average ratio of toxic chemical emissions to revenues during 1999–2000 was above or below its industry median (50th percentile) value. We also tried categorizing firms based on whether they were above or below their industry's 40th percentile. We used the 60th percentile as a second alternative threshold. Results obtained using the alternative thresholds were similar to our main results.

We also tested several modifications to how we measured and categorized KLD ratings. In one model, instead of just two categories of newly rated firms, poor and good or mixed, we employed four—(1) only concerns (poor), (2) only strengths (good), (3) mix of strengths and concerns (mixed), and (4) no strengths and no concerns (null)—and interacted each of these with a dummy coded '1' for years in which a firm had been rated by KLD. The results provide (by construction) the identical coefficient estimate for firms that received a poor KLD rating (our focal firms), but also reveal that firms rated good or mixed performed no differently than firms that were never rated (i.e., the control group). These results are presented in Column 1 of Table A1 in the Appendix.

In a second model, we created two numeric measures based on firms' initial KLD ratings: (1) the sum of strengths and (2) the sum of concerns. We interacted each of these with a dummy coded '1' for years in which a firm had been rated by KLD, and '0' otherwise. The results, which indicate that firms with more concerns in their initial KLD rating reduced their emissions significantly more (Column 2 of Table A1), are consistent with our main findings.[11]

Finally, we interacted a dummy coded '1' in years when a firm had been rated by KLD (and '0' otherwise) with a series of dummies that reflected whether a firm's initial KLD concern ratings included compliance concerns (*hazardous waste* and *regulatory problems*), emissions concerns (*substantial emissions* and *climate change*), and other concerns (*agricultural chemical, ozone-depleting substance*, and *other concerns*). The results of this model indicate that firms initially identified by KLD as having concerns related to compliance or emissions subsequently improved their environmental performance (Column 3 of Table A1), which is consistent with our main results. Interestingly, we find no evidence that such improvements were associated with initially being identified as having other concerns (or any strengths), which suggests that firms identified as having potential compliance or emissions issues but not necessarily suffering from other environmental problems are more likely to reduce emissions in response to poor ratings. Estimating this model predicting the annual number of penalties (explained below), using a conditional fixed effects negative binomial model, yielded similar results (Column 4 of Table A1).

We also tested the robustness of our results to several changes to our sample. We reestimated the models that tested Hypothesis 1 on the slightly smaller samples used to test Hypothesis 3 by omitting firms that, in the absence of emissions and revenue data for 1999 or 2000, could not be classified as being more or less environmentally efficient. We also reestimated our main models on the subsample that excluded firms with initial ratings that contained neither environmental strengths nor environmental concerns, as such 'null ratings' might be due to KLD being unable to acquire the needed information rather than making an informed determination that these firms actually had no strengths and no concerns. Finally, we reestimated these models excluding firms never rated during the sample period. All firms in the resulting subsample went through the transition from being unrated to being rated during the sample period. Estimating our models on each of these alternative samples provided additional statistically significant support for our three hypotheses.

There is also the concern that our results might be driven by mean reversion, as in the following hypothetical scenario. Suppose that annual emission level changes are largely a random process such that firms that exhibit unusually high emission levels, and are consequently rated poor by KLD, subsequently exhibit reduced emissions due simply to random fluctuations. Mean reversion would predict that firms rated favorably by KLD would subsequently exhibit higher emissions. Empirical evidence from prior research, together with the robustness tests we conducted, lead us to believe

[11] We did not create a single 'net' score by aggregating both strengths and concerns because prior research has demonstrated that KLD strengths and KLD concerns represent distinct constructs (Mattingly and Berman, 2006).

DOI: 10.1002/smj

that our results are not being driven by mean reversion. Observations of *higher* subsequent emission levels at firms consistently rated poor by KLD (Chatterji *et al.*, 2009) are contrary to what would be predicted by a mean reversion hypothesis.[12] Furthermore, we performed a robustness test that compared newly rated firms to an alternative comparison group, S&P 500 Index member firms that received KLD ratings throughout our sample period. The results indicated that newly rated firms initially rated poor by KLD subsequently improved more than always-rated firms that had been rated poor during the pre-rating period. We found no such difference between the comparable groups that were initially rated mixed or good. Because these models compare treatment and control firms with the same initial or pre-period ratings, the subsequent performance differences among the firms initially rated poor are not driven by mean reversion. These robustness tests are described more fully in the Appendix, where their results are displayed in Column 1 of Tables A2–A4.

Finally, we considered an alternative measure of environmental performance based on environmental regulatory compliance (Helland, 1998; Sharma, 2000; Short and Toffel, 2008). We obtained the *annual number of penalties* each firm accrued for violating regulations associated with all nine major U.S. federal environmental regulations included in the CEPD database.[13] This being a count dependent variable, we estimated these models using a conditional fixed effects negative binomial specification. Results were similar to those obtained in our main analysis. That firms initially rated poor subsequently accrued significantly fewer penalties than both never-rated firms and firms initially rated mixed or good provides additional support for Hypothesis 1. Similarly, less environmentally efficient firms initially rated poor subsequently accrued significantly fewer penalties than other less environmentally efficient firms, but the absence of this pattern within the more environmentally efficient group, lends further support to Hypothesis 3. The finding, in contrast to our main results, that firms rated poor by KLD experienced significant reductions in penalty rates in both highly *and* less intensively regulated industries does not support Hypothesis 2.[14] The results of these penalty models are displayed in Column 2 of Tables A2–A4 in the Appendix.

Overall, the results of these tests demonstrate that our main results are robust to a variety of plausible alternative independent measures and several alternative samples, and two of our three main results are robust to an alternative dependent variable.

DISCUSSION

We find that firms that initially received poor KLD ratings subsequently improved their environmental performance more than other firms, and that this difference was driven by firms in highly regulated industries and by firms with more low-cost opportunities to exploit. To our knowledge, our work represents one of the first efforts to develop a theory that predicts which firms will change their performance in response to ratings and introduce important contingencies that influence firms' responses. Our research design is based on an exogenous change in the rating status of some of the firms in our sample. This empirical strategy moves us closer to the ideal of a randomized experiment, still rare in management research.

Contributions

The central contribution of our work is to a nascent organizational literature that examines the impact of ratings on rated organizations. We argue that organizations will improve their performance in response to poor ratings to mitigate the threat of stakeholder sanctions. In contrast to our theoretical predictions and results, Espeland and Sauder

[12] Results of prior research are based on assessing the ongoing dynamic relationship between annual ratings and subsequent annual emissions over a 13-year period (Chatterji *et al.*, 2009). Our current analysis examines the relationship between ratings and subsequent emissions for *newly* rated firms, a small fraction of the firms rated by KLD.

[13] These include: the Atomic Energy Act; Clean Air Act; Clean Water Act; Endangered Species Act; Federal Insecticide, Fungicide, and Rodenticide Act; Mine Safety and Health Act; Resource Conservation and Recovery Act; Safe Drinking Water Act; and Toxic Substances Control Act.

[14] When comparing the results of these negative binomial models to those obtained from the OLS models in the main analysis, an important caveat is that the conditional fixed effects negative binomial models are estimated based on data only for firms in the sample that exhibited variation in the number of penalties during the sample period. In fact, fewer than half the firms in the sample exhibited such variation, which might account for some of the differences in results between the emissions and penalty regressions.

(2007) found that low rankings accelerated the decline of law schools by eroding their ability to attract high quality applicants and raise funds. Our results display the opposite pattern: poorly rated firms subsequently improved their performance. Further research is needed to understand the circumstances under which poor ratings will motivate or enervate organizations.

Our research goes beyond prior approaches described in this literature by hypothesizing that two important contingencies—firms' efficiency levels and regulatory environments—moderate how firms respond to ratings. Future research could explore other contingencies such as differences across industries, competitive environments, or CEO characteristics.

Our research also contributes to instrumental stakeholder theory, which proposes that responsiveness to stakeholders positively affects firms' performance (Jones, 1995). More recent theoretical and empirical developments in this domain suggest that firms' responsiveness to stakeholders depends on key characteristics of the *stakeholders* who issue requests and the nature of those *requests* (e.g., power, legitimacy, urgency) (Eesley and Lenox, 2006; Mitchell *et al.*, 1997). Our work makes two contributions to this literature. First, we posit that firms' responsiveness to ratings that target stakeholders in one domain (e.g., consumers and investors) can be moderated by the extent to which firms are threatened by a stakeholder in another domain (e.g., the government). Second, we argue that firms facing lower cost opportunities will be more likely to respond to low ratings. In doing so, we supplement the nascent literature focused on how firm characteristics influence responsiveness to information disclosure (Konar and Cohen, 1997; Lenox and Eesley, 2009).

Our work also offers insights that can inform institutional theory. Poor ratings might, indeed, threaten firms' legitimacy (Hunter and Bansal, 2007), but our study reveals that other factors including efficiency and regulatory context influence organizational responses. Whereas much of the theoretical work on institutional theory focuses on deterministic constraints imposed by institutional forces, our findings suggest that the influence of such forces can be supported or constrained by organizations' strategic choices (Child, 1972). Finally, whereas the economics literature has long considered the role of information asymmetry and potential of information disclosure programs to influence behavior, our work demonstrates considerable variation in incentives for firms to respond to information disclosure, shaped by firm-level characteristics and the threat of regulation.

While our theory focused on firms' responses to negative ratings, further research is needed to understand how organizations respond to positive ratings. Positive ratings might elevate performance by enabling organizations to lower their cost of capital and attract higher quality talent and more prestigious supply chain partners. Positive ratings might also reduce stakeholder scrutiny or managerial attention to the rated issue, either of which could lead some organizations to rest on their laurels and risk a subsequent decline in performance.

Finally, because understanding how firms respond to ratings is central to the study of firm strategy, further research could examine this issue in other empirical contexts. For example, do firms that rate poorly on diversity subsequently place more women or minorities on their boards of directors? Do poor ratings on *Angie's List* lead home contractors to improve customer service or lower their prices? Do producers of coffee makers rated high by *Consumer Reports* expand their production runs or raise their prices?

Policy implications

We have found changes in organizational performance to be associated with ratings issued by an independent rating agency. Although this study is, to our knowledge, the first to identify this effect with independent, non-governmental rating agencies, our results are consistent with the findings of prior research that examined the effects of government information disclosure programs on firm behavior (e.g., Greenstone, Oyer, and Vissing-Jorgensen, 2006; Jin and Leslie, 2003). Our findings are most similar to those of a study that found that companies rated by a government program in Indonesia as having the worst environmental performance with respect to water pollution subsequently made the greatest improvements (Blackman *et al.*, 2004).

Our study provides empirical support for the key assumption underlying 'information-based regulation' that focuses on information disclosure rather than behavior control. Various forms of information-based regulation—whether requirements that fast food restaurants include nutritional information on menus or that industrial facilities publicly

DOI: 10.1002/smj

disclose toxic chemical emissions and, as currently proposed, greenhouse gases—are predicated on the notion that responses of consumers, investors, or other important stakeholders will motivate firms to improve their performance. Our results supplement a growing body of empirical research on government programs that mandate greater transparency. Evidence suggests that some of these government programs, much like the third-party program we study, prompt firms to improve their environmental performance (Blackman *et al.*, 2004; Fung, Graham, and Weil, 2007; Scorse, 2007). Our theory and findings can also help inform emerging theoretical research that is exploring the circumstances under which government mandatory information disclosure programs are particularly likely to achieve their policy objectives (Cohen and Santhakumar, 2007; Fung *et al.*, 2007).

Although our study is the first to examine how companies respond to third party ratings, distinctions between government and third party efforts are not necessarily hard and fast. KLD's environmental ratings are based in part on historical government data extracted from government databases (e.g., regulatory compliance records, number of Superfund sites), and much of its ability to predict environmental outcomes derives from its aggregation of these data (Chatterji *et al.*, 2009). This point highlights an opportunity for policy makers to partner with other stakeholder groups: governments can exercise their coercive power to gather data from companies while stakeholder groups can focus on communicating the information to the public. Although some firms might respond to nonmarket mechanisms alone, others might be more susceptible to pressure from such collaborations between government and third party organizations. Future research should explore these kinds of partnerships and assess their effectiveness at influencing different kinds of organizations.

Interestingly, examples can be found of nongovernmental entities already communicating data to the public with little involvement by government. Consider that the data solicited annually by the EPA from tens of thousands of facilities on the use and emissions of more than 600 toxic chemicals languishes on two fairly obscure EPA Web sites (www.epa.gov/tri and www.epa.gov/enviro). To make these data more visible and useful, Environmental Defense and The Right-to-Know Network each created user-friendly Web portals (www.scorecard.org and www.rtknet.org, respectively), a team of academics created a Google Map mash-up of the data (www.mapecos.org; see Walker (2008)), and the Investor Responsibility Research Center aggregated the factory-level data to the parent companies to create the CEPD. In this spirit, *Wikinomics* author Anthony Williams foresees a future in which non-governmental organizations and other sectors create user-friendly Web portals to aggregate data from government and other sources, transform it into information of public value, and distribute it (Williams, 2007).

The results of our study have policy implications for boosting the effectiveness of government information disclosure programs. Government agencies striving to leverage mandatory information disclosure programs to improve the environmental performance of laggard enterprises might take recourse to information-based incentives such as 'shaming,' a strategy that might be particularly effective in highly regulated industries.

Regulators might accompany the 'stick' of information-based incentives with a 'carrot' in the form of helping firms identify opportunities for low-cost improvements. In practical terms, policy makers can promote change by lowering the cost of investments in environmental performance improvements, such as providing technical assistance or subsidies to facilitate knowledge transfer to or between firms. Government technical assistance programs (O'Rourke and Lee, 2004) might be ideally suited to help companies, especially those not yet shamed by an external rating, identify opportunities for low-cost improvements. Deploying scarce technical assistance resources to help less environmentally efficient firms improve their performance could yield much greater aggregate performance improvement than dispensing such resources on a first-come-first-served basis. Alternatively, governments might promote technical assistance through subsidies, as the governments of Pakistan and Singapore have done in subsidizing the training associated with companies' adoption of international environmental and labor standards, and the EPA has done in sponsoring its 'National Environmental Partnership Summit' to facilitate the sharing of best practices among industry participants. These mechanisms will be especially pertinent in technology-intensive industries in which much knowledge is tacit and difficult to transfer.

The insights yielded by this study can also be broadly applied in other policy arenas, notably

DOI: 10.1002/smj

in education policy. For example, the No Child Left Behind Act, a U.S. law passed in 2001, uses shaming mechanisms to identify failing schools, arguably without providing the necessary resources for improvement (Linn, Baker, and Betebenner, 2002). Our work suggests that this kind of policy would be more effective if failing schools were provided with increased funding to identify low-cost opportunities to raise student achievement.

Limitations

We acknowledge a number of limitations to our study. For example, since our dataset ends in 2004, we are unable to determine whether the firms in which we observed improvements maintained those improvements. Future work could analyze organizational responses to ratings over longer periods of time. In addition, our empirical analysis employs firm-level fixed effects to examine performance differences within firms over time. These fixed effects control for any influence of managerial effectiveness that might also affect environmental performance, environmental efficiency, or the KLD ratings—to the extent that such influence within firms remains constant over time throughout the sample period. That said, it is possible that during our sample period managerial effectiveness in some firms independently improved (or worsened) unobserved, and that these changes affected those firms' environmental performance, environmental efficiency, or KLD ratings. In that case, our results could suffer omitted variable bias. To affect the inferences from our analysis, however, this would have had to occur disproportionately among the newly rated (treatment) group or the never-rated (control) group. We have no reason to suspect this concern to seriously bias our results, but nonetheless acknowledge it as a possibility. Another limitation of our analysis is that we were unable to obtain data to control for the age of companies' facilities and their environmental control technologies. Future research could explore whether and how these factors might influence organizations' environmental performance and ratings.

Although we have relied on an exogenous shock and employed a quasi-control group, we cannot be certain that firms are responding directly to these ratings and not to other forces in the political, economic, or social environment that might be related to the ratings. Understanding how and why firms respond differently to negative ratings is an important avenue for future research.

A related limitation of our study is that the average firm size differs between our treatment and quasi-control groups, which provides for a less than ideal comparison. We control for changes in firm size by including log employees, log sales, log assets, and log number of TRI-reporting facilities in all of our models, and we control for each firm's average size by including firm fixed effects. We also note that we compare firms that became rated during the sample period (our treatment group) to two different control groups. In the main analysis, we compare them to firms that were never rated during the sample period, a control group that features companies that are smaller on average than those in the treatment group. By contrast, in our robustness tests we compare our treatment group to firms that were already rated at the beginning of our sample period, and the firms in that control group are larger on average than those in than the treatment group. Because we find similar results when comparing our treatment group to either of these control groups, we think it is unlikely that our results are driven by differences in average firm size between the treatment and control groups.

Another potential limitation relates to the TRI data on which we rely to create our emissions-based measure of environmental performance. While facilities in specified industries with employment and emission levels beyond particular thresholds are legally required to report these data to the EPA, several concerns have been raised about TRI data including lax regulatory verification of these self-reported data, potentially confusing changes in the regulator's instructions, that the data include only a subset of pollutants, and that some TRI data are estimated rather than measured (Frey and Small, 2003; Gerde and Logsdon, 2001; Toffel and Marshall, 2004). Nonetheless, TRI data are among the most widely used by academics to measure corporate environmental performance, and our additional use of regulatory penalties as an alternative measure of environmental performance indicates that our conclusions are robust to such concerns about TRI data.

Lastly, while KLD ratings data have been widely used in studies of corporate social responsibility and socially responsible investing (Berman *et al.*, 1999; Margolis and Walsh, 2003), KLD data have limitations. KLD environmental ratings focus on only a subset of environmental issues,

DOI: 10.1002/smj

and implicitly give equal weight to each of the issues. During our sample period, KLD strengths and concerns were a series of binary measures, which homogenizes issues that in reality can differ significantly in magnitude. In addition, KLD data are largely based on the performance of the rated firms' U.S. subsidiaries rather than their worldwide operations. Finally, it is possible that information asymmetries between raters and the rated can lead to measurement error. Despite these concerns, prior research has found empirical support for the construct validity (Sharfman, 1996) and predictive validity (Chatterji *et al.*, 2009) of KLD ratings, and we believe these drawbacks about KLD data are unlikely to have biased our results.

CONCLUSION

Company ratings and rankings have a long history and continue to proliferate. Our paper is the first to provide theoretical guidance toward understanding how firms change their performance in response to third-party ratings. We contribute to existing theories by providing a more nuanced view of how firms do so. Future research should investigate whether other kinds of independent raters and market intermediaries exert a similar impact. Worthy candidates for such research include Moody's and S&P as well as agencies that consolidate user-based ratings such as *Zagat's* and *Angie's List*. Future research could also examine the relationships we explored in other domains including education (e.g., how public schools respond to ratings from the No Child Left Behind program) and product quality (e.g., how manufacturers respond to *Consumer Reports* ratings). More broadly, we hope our work is part of a nascent literature that explores how ratings influence both organizations and their stakeholders.

ACKNOWLEDGEMENTS

We thank Amy Edmondson, Andrew King, Karim Lakhani, Michael Lenox, David Levine, Jon Kolstad, Joshua Margolis, Chris Marquis, Will Mitchell, Lamar Pierce, Christopher Rider, Jason Scorse, Jason Snyder, Anita Tucker, Bennet Zelner, two anonymous reviewers, and seminar participants from Harvard University, Duke University, and the Strategy and Business Environment Conference for their helpful comments and suggestions. We gratefully acknowledge Joe Henzlik of Institutional Shareholder Services and Jennifer Hodge and Michael Langlais of RiskMetrics Group for providing and helping us to understand the Corporate Environmental Profiles Directory dataset. We also thank the Division of Research at the Harvard Business School for financial support

REFERENCES

Akerlof GA. 1970. The market for 'lemons': quality uncertainty and the market mechanism. *Quarterly Journal of Economics* **84**(3): 488–500.

Ayres RU. 1995. Achieving eco-efficiency in business. Report of the WBCSD Second Antwerp Eco-Efficiency Workshop. Paper presented at the Geneva World Business Council for Sustainable Development. (14–15 March).

Bansal P, Clelland I. 2004. Talking trash: legitimacy, impression management, and unsystematic risk in the context of the natural environment. *Academy of Management Journal* **47**(1): 93–103.

Barnett M, Salomon R. 2006. Beyond dichotomy: the curvilinear relationship between social responsibility and financial performance. *Strategic Management Journal* **27**(11): 1101–1122.

Baue W. 2003. TIAA-CREF adds positive screens and replicates Russell 3000 performance. Available at: http://www.socialfunds.com/news/article.cgi/1128.html (16 July, 2008).

Becchetti L, Ciciretti R, Hasan I. 2007. Corporate social responsibility and shareholder's value: an event study analysis. Working paper 2008-06, Federal Reserve Bank of Atlanta, Atlanta, GA.

Bennear LS, Olmstead SM. 2008. The impacts of the 'right to know': information disclosure and the violation of drinking water standards. *Journal of Environmental Economics and Management* **56**(2): 117–130.

Berman SL, Wicks AC, Kotha S, Jones TM. 1999. Does stakeholder orientation matter? The relationship between stakeholder management models and firm financial performance. *Academy of Management Journal* **42**(5): 488–506.

Blackman A, Afsah S, Ratunanda D. 2004. How do public disclosure pollution control programs work? Evidence from Indonesia. *Human Ecology Review* **11**(3): 235–246.

Chatterji AK, Levine DI. 2006. Breaking down the wall of codes: evaluating non-financial performance measurement. *California Management Review* **48**(2): 29–51.

Chatterji AK, Levine DI, Toffel MW. 2009. How well do social ratings actually measure corporate social responsibility? *Journal of Economics and Management Strategy* **18**(1): 125–169.

Child J. 1972. Organizational structure, environment and performance: the role of strategic choice. *Sociology* **6**(1): 114–127.

Cho CH, Patten DM. 2007. The role of environmental disclosures as tools of legitimacy: a research note. *Accounting Organizations and Society* **32**(7/8): 639–647.

Christmann P. 2000. Effects of 'best practices' of environmental management on cost advantage: the role of complementary assets. *Academy of Management Journal* **43**(4): 663–680.

Coglianese C, Nash J (eds.). 2001. *Regulating from the Inside: Can Environmental Management Systems Achieve Policy Goals?* Resources for the Future: Washington, D.C.

Cohen MA, Santhakumar V. 2007. Information disclosure as environmental regulation: a theoretical analysis. *Environmental and Resource Economics* **37**(3): 599–620.

Corbett CJ, Montes-Sancho MJ, Kirsch DA. 2005. The financial impact of ISO 9000 certification in the United States: an empirical analysis. *Management Science* **51**(7): 1046–1059.

Curran MM, Moran D. 2007. Impact of FTSE4good index on firm price: an event study. *Journal of Environmental Management* **82**(4): 529–537.

Cutler DM, Huckman RS, Landrum MB. 2004. The role of information in medical markets: an analysis of publicly reported outcomes in cardiac surgery. *AEA Papers and Proceedings* **94**(2): 342–346.

Darnall N, Edwards D. 2006. Predicting the cost of environmental management system adoption: the role of capabilities, resources, and ownership structure. *Strategic Management Journal* **27**(4): 301–320.

Deckop JR, Merriman KK, Gupta S. 2006. The effects of CEO pay structure on corporate social performance. *Journal of Management* **32**(3): 329–342.

Delmas MA. 2002. The diffusion of environmental standards in Europe and in the United States. *Policy Sciences* **35**(1): 91–119.

Delmas MA, Russo MV, Montes-Sancho MJ. 2007. Deregulation and environmental differentiation in the electric utility industry. *Strategic Management Journal* **28**(2): 189–209.

Delmas MA, Toffel MW. 2008. Organizational responses to environmental demands: opening the black box. *Strategic Management Journal* **29**(10): 1027–1055.

Eesley C, Lenox MJ. 2006. Firm responses to secondary stakeholder action. *Strategic Management Journal* **27**(8): 765–781.

Eesley CE, Lenox MJ. 2009. Private environmental activism and the selection and response of firm targets. *Journal of Economics & Management Strategy* **18**(1): 45–73.

Elsbach KD, Kramer RM. 1996. Members' responses to organizational identity threats: encountering and countering the business week rankings. *Administrative Science Quarterly* **41**(3): 442–476.

Espeland WN, Sauder M. 2007. Rankings and reactivity: how public measures recreate social worlds. *American Journal of Sociology* **113**(1): 1–40.

Fombrun CJ. 2007. List of lists: a compilation of international corporate reputation ratings. *Corporate Reputation Review* **10**(2): 144–153.

Fortune. 2008. 100 Best companies to work for. Available at: http://money.cnn.com/magazines/fortune/bestcompanies/2008/ (16 July, 2008).

Frey HC, Small MJ. 2003. Integrated environmental assessment, part 1: estimating emissions. *Journal of Industrial Ecology* **7**(1): 9–11.

Fung A, Graham M, Weil D. 2007. *Full Disclosure: The Promise and Perils of Transparency*. Cambridge University Press: Cambridge, UK.

Gerde VW, Logsdon JM. 2001. Measuring environmental performance: use of the toxics release inventory (TRI) and other U.S. environmental databases. *Business Strategy & the Environment* **10**(5): 269–285.

Goodstein JD. 1994. Institutional pressures and strategic responsiveness: employer involvement in work-family issues. *Academy of Management Journal* **37**(2): 350–382.

Graham M. 2000. Regulation by shaming. *Atlantic Monthly* **285**(4): 36–40.

Graham M, Miller C. 2001. Disclosure of toxic releases in the United States. *Environment* **43**(8): 8–20.

Greenstone M, Oyer P, Vissing-Jorgensen A. 2006. Mandated disclosure, stock returns, and the 1964 securities acts amendments. *Quarterly Journal of Economics* **121**(2): 399–460.

Gunningham N, Kagan RA, Thornton D. 2004. Social license and environmental protection: why businesses go beyond compliance. *Law & Social Inquiry* **29**(2): 307–341.

Hamilton JT. 1995. Pollution as news: media and stock market reactions to the toxics release inventory data. *Journal of Environmental Economics and Management* **28**(1): 98–113.

Hannan EL, Kilburn H, Racz M, Shields E, Chassin M R. 1994. Improving the outcomes of coronary artery bypass surgery in New York State. *Journal of the American Medical Association* **271**(10): 761–766.

Hart SL. 1995. A natural resource-based view of the firm. *Academy of Management Review* **20**(4): 986–1014.

Hart SL, Ahuja G. 1996. Does it pay to be green? An empirical examination of the relationship between emission reduction and firm performance. *Business Strategy & the Environment* **5**(1): 30–37.

Helland E. 1998. The enforcement of pollution control laws: inspections, violations, and self-reporting. *Review of Economics and Statistics* **80**(1): 141–153.

HRC. 2008. Human Rights Campaign Best Places to Work. Available at: http://www.hrc.org/issues/workplace/best_places_to_work_2008.asp (22 July, 2008).

Hunter T, Bansal P. 2007. How standard is standardized MNC global environmental communication? *Journal of Business Ethics* **71**(2): 135–147.

Jin GZ, Leslie P. 2003. The effect of information on product quality: evidence from restaurant hygiene grade cards. *Quarterly Journal of Economics* **118**(2): 409–451.

Jin GZ, Leslie P. 2009. Reputational incentives for restaurant hygiene. *American Economic Journal: Microeconomics* **1**(1): 237–267.

Jones TM. 1995. Instrumental stakeholder theory: a synthesis of ethics and economics. *Academy of Management Review* **20**(2): 404–442.

Kassinis G, Vafeas N. 2006. Stakeholder pressures and environmental performance. *Academy of Management Journal* **49**(1): 145–159.

Kempf A, Osthoff P. 2007. The effect of socially responsible investing on portfolio performance. *European Financial Management* **13**(5): 908–922.

Khanna M, Quimio WH, Bojilova D. 1998. Toxics release inventory: a policy tool for environmental protection. *Journal of Environmental Economics and Management* **36**(3): 243–299.

King AA, Lenox MJ. 2000. Industry self-regulation without sanctions: the chemical industry's responsible care program. *Academy of Management Journal* **43**(4): 698–716.

King AA, Lenox MJ. 2002. Exploring the locus of profitable pollution reduction. *Management Science* **48**(2): 289–299.

King AA, Lenox MJ, Terlaak A. 2005. The strategic use of decentralized institutions: exploring certification with the ISO 14001 management standard. *Academy of Management Journal* **48**(6): 1091–1106.

Klassen RD, Whybark DC. 1999. The impact of environmental technologies on manufacturing performance. *Academy of Management Journal* **42**(6): 599–615.

KLD.com. 2006. KLD Research & Analytics, Inc. Available at: http://www.kld.com (11 September, 2006).

Konar S, Cohen MA. 1997. Information as regulation: the effect of community right to know laws on toxic emissions. *Journal of Environmental Economics and Management* **32**(1): 109–124.

Lenox MJ, Eesley CE. 2009. Private environmental activism and the selection and response of firm targets. *Journal of Economics and Management Strategy* **18**(1): 45–73.

Leone RA. 1981. *Government Regulation of Business: Developing the Managerial Perspective.* Harvard Business School Press: Boston, MA.

Leone RA. 1986. *Who Profits: Winners, Losers, and Government Regulation.* Basic Books: New York.

Levi MD, Nault BR. 2004. Converting technology to mitigate environmental damage. *Management Science* **50**(8): 1015–1030.

Linn R, Baker E, Betebenner D. 2002. Accountability systems: implications of requirements of the No Child Left Behind Act of 2001. *Educational Researcher* **31**(6): 3–16.

Lyon TP, Maxwell JW. 2001. Voluntary approaches to environmental regulation: a survey. In *Economic Institutions and Environmental Policy,* Franzini M, Nicita A (eds). Ashgate Publishing: Aldershot, UK; 75–120.

Margolis JD, Walsh JP. 2003. Misery loves companies: rethinking social initiatives by business. *Administrative Science Quarterly* **48**(2): 268–305.

Mattingly JE, Berman SL. 2006. Measurement of corporate social action: discovering taxonomy in Kinder Lydenberg Domini ratings data. *Business and Society* **45**(1): 20–46.

Mitchell RK, Agle BR, Wood DJ. 1997. Toward a theory of stakeholder identification and salience: defining the principle of who and what really counts. *Academy of Management Review* **22**(4): 853–886.

Moodys.com. 2008. About Moodys. http://www.moodys.com/moodys/cust/AboutMoodys/ (17 July, 2008).

Norton F. 2007. 'Green' investments growing in popularity. *Philadelphia Inquirer*. 21 August.

Oberholzer-Gee F, Mitsunari M. 2006. Information regulation: do the victims of externalities pay attention? *Journal of Regulatory Economics* **30**(2): 141–158.

O'Rourke D, Lee E. 2004. Mandatory planning for environmental innovation: evaluating regulatory mechanisms for toxics use reduction. *Journal of Environmental Planning and Management* **47**(2): 181–200.

Peterson ED, DeLong ER, Jollis JG, Muhlbaier LH, Mark DB. 1998. The effects of New York's bypass surgery provider profiling on access to care and patient outcomes in the elderly. *Journal of the American College of Cardiology* **32**(4): 993–999.

Ramus CA, Killmer ABC. 2007. Corporate greening through prosocial extrarole behaviours—a conceptual framework for employee motivation. *Business Strategy & the Environment* **16**(8): 554–570.

Ramus CA, Steger U. 2000. The roles of supervisory support behaviors and environmental policy in employee 'ecoinitiatives' at leading-edge European companies. *Academy of Management Journal* **43**(4): 605–626.

Rao H. 1994. The social construction of reputation: certification contests, legitimation, and the survival of organizations in the American automobile industry: 1895–1912. *Strategic Management Journal*, Winter Special Issue **15**: 29–44.

Reinhardt FL. 1998. Environmental product differentiation: implications for corporate strategy. *California Management Review* **40**(4): 43–74.

Rock MT. 2003. Public disclosure of the sweatshop practices of American multinational garment/shoe makers/retailers: impacts on their stock prices. *Competition & Change* **7**(1): 23–38.

Russo MV, Fouts PA. 1997. A resource-based perspective on corporate environmental performance and profitability. *Academy of Management Journal* **40**(3): 534–559.

Russo MV, Harrison NS. 2005. Organizational design and environmental performance: clues from the electronics industry. *Academy of Management Journal* **48**(4): 582–593.

Savitz AW, Weber K. 2007. The sustainability sweet spot. *Environmental Quality Management* **17**(2): 17–28.

Scorse J. 2007. Does being a 'top 10' worst polluter affect facility environmental releases? Evidence from the U.S. Toxic Release Inventory. Working paper, Monterey Institute of International Studies, Monterey, CA.

Scott WR. 1987. The adolescence of institutional theory. *Administrative Science Quarterly* **32**(4): 493–511.

Sen S, Bhattacharya C. 2001. Does doing good always lead to doing better? Consumer reactions to corporate social responsibility. *Journal of Marketing Research* **38**(2): 225–243.

Sharfman M. 1996. The construct validity of the Kinder, Lydenberg & Domini social performance ratings data. *Journal of Business Ethics* **15**(3): 287–296.

Sharma S. 2000. Managerial interpretations and organizational context as predictors of corporate choice of environmental strategy. *Academy of Management Journal* **43**(4): 681–694.

Sharma S, Vredenburg H. 1998. Proactive corporate environmental strategy and the development of competitively valuable organizational capabilities. *Strategic Management Journal* **19**(8): 729–753.

Short JL, Toffel MW. 2008. Coerced confessions: self-policing in the shadow of the regulator. *Journal of Law, Economics, & Organization* **24**(1): 45–71.

Starik M, Marcus AA. 2000. Introduction to the special research forum on the management of organizations in the natural environment: a field emerging from multiple paths, with many challenges ahead. *Academy of Management Journal* **43**(4): 539–546.

Stephan M. 2002. Environmental information disclosure programs: they work but why? *Social Science Quarterly* **83**(1): 190–205.

Takeda F, Tomozawa T. 2008. A change in market responses to the environmental management ranking in Japan. *Ecological Economics* **67**(3): 465–472.

Terlaak A. 2007. Order without law? The role of certified management standards in shaping socially desired firm behaviors. *Academy of Management Review* **32**(3): 968–985.

TIAA-CREF. 2008. Institutional class mutual funds: social choice equity. http://www.tiaacref.org/performance/institutional/profiles/1834.html (8 May, 2008).

TIAA-CREF. 2009. TIAA-CREF funds: retail class: social choice equity. http://www.tiaa-cref.org/performance/mutual_funds/profiles/0059.html (12 December, 2009).

Toffel MW, Marshall JD. 2004. Improving environmental performance assessment: comparative analysis of weighting methods used to evaluate chemical release inventories. *Journal of Industrial Ecology* **8**(1–2): 143–172.

Tomkins R. 2001. When caring is a good investment. *Financial Times* London. 5 October.

Vogel D. 2005. *The Market for Virtue: The Potential and Limits of Corporate Social Responsibility.* Brookings Institution Press: Washington, D.C.

von Hippel E. 1988. *The Sources of Innovation.* Oxford University Press: New York.

Waddock S. 2003. Myths and realities of social investing. *Organization and Environment* **16**(3): 369–380.

Waddock SA, Graves SB. 1997. The corporate social performance–financial performance link. *Strategic Management Journal* **18**(4): 303–317.

Walker P. 2008. Mapping out the environment. http://www.cnn.com/2007/BUSINESS/12/11/execed.mapecos/ (12 March, 2008).

Weil D, Fung A, Graham M, Fagotto E. 2006. The effectiveness of regulatory disclosure policies. *Journal of Policy Analysis and Management* **25**(1): 155–181.

Wilbert C. 2006. Social responsibility of Coke questioned. *Atlanta Journal Constitution.* 19 July.

Williams AD. 2007. *Gov 2.0: The Power of Collaborative Communities Connecting Communities: Using Information to Drive Change*. Brookings Institution: Washington, D.C.

APPENDIX

EXHIBIT A1: ROBUSTNESS TESTS REGARDING MEAN REVERSION

Whereas our main analysis compares newly rated firms to never-rated firms, here we compare newly rated firms to S&P 500 member firms that had KLD ratings throughout the sample period. We categorize these always-rated firms into two groups based on whether their KLD ratings during 1999–2000 (the pre-period of our main analysis) included (1) only of KLD concerns, or (2) no KLD concerns or also KLD strengths. This categorization mirrors our classification of the newly rated firms (based on their initial KLD ratings). We refer to the group of (a) newly rated firms with poor initial ratings and (b) always-rated firms rated poor during the pre-period as the *concerns ex ante comparison group* (and create a dummy variable denoting membership in this group). We refer to the group of (a) newly rated firms with mixed or good initial ratings and (b) always-rated firms rated mixed or good during the pre-period as the *mixed/good ex ante comparison group*. We compare the performance of the newly rated firms to that of the always-rated firms *within* each *ex ante comparison group*. Because all members within each group received similar initial ratings, subsequent performance differences between the newly rated and always-rated firms within these comparison groups cannot be due to mean reversion. We modify our three main models by adding interaction terms between the concerns *ex ante* comparison group dummy and all other variables.[15] The results indicate that newly rated firms with poor initial KLD ratings subsequently improved compared to always-rated firms that had poor ratings during the pre-period. We found no evidence that newly rated firms with mixed or good initial KLD ratings subsequently outperformed the always-rated firms that had mixed or good ratings during the pre-period. These results, which refute concerns that our main results are a function of mean reversion, are displayed in Column 1 of Tables A2–A4.

[15] Including interaction terms between a mixed/good comparison group dummy and all other variables would yield the same result because the two comparison group dummies are coded exactly opposite.

DOI: 10.1002/smj

Table A1. Results of robustness tests

	Dependent variable:	1 Log emissions OLS coefficients	2 Log emissions OLS coefficients	3 Log emissions OLS coefficients	4 Number of regulatory penalties Incident rate ratios
	KLD rated × initial rating poor	−0.656***			
		[0.254]			
	KLD rated × initial rating good	1.199			
		[0.729]			
	KLD rated × initial rating mixed	0.039			
		[0.670]			
	KLD rated × initial rating null	0.298			
		[0.172]*			
(A)	KLD rated × sum of initial KLD concerns		−0.610***		
			[0.145]		
(B)	KLD rated × sum of initial KLD strengths		0.721*		
			[0.419]		
(C)	KLD rated × initial rating includes compliance-related concerns[a]			−0.706**	0.428***
				[0.353]	[0.135]
(D)	KLD rated × initial rating includes emissions-related concerns[b]			−0.913***	0.701
				[0.293]	[0.185]
	KLD rated × initial rating includes other concerns[c]			0.044	0.799
				[0.461]	[0.295]
(E)	KLD rated × initial rating includes at least one strength			0.921*	1.767*
				[0.515]	[0.595]
	Log employees	−0.313	−0.274	−0.262	1.402
		[0.241]	[0.241]	[0.241]	[0.225]
	Log sales	0.718***	0.679***	0.670***	1.005
		[0.234]	[0.234]	[0.234]	[0.18]
	Log assets	−0.181	−0.121	−0.129	0.732
		[0.265]	[0.264]	0.264	[0.148]
	Log number of TRI-reporting facilities	1.778***	1.780***	1.797***	0.956
		[0.142]	[0.142]	0.142	[0.081]
	Year dummies (2000–2004)	Included	Included	Included	Included
	Firm-level fixed effects	Included	Included	Included	Included[d]
	Observations (firm-years)	2412	2412	2412	1122
	Firms	598	598	598	228
	R-squared (within)	0.16	0.17	0.16	
	Wald test: coefficient on (A) = (B)?		8.13***		
	Wald test: coefficient on (D) = (E)?			9.62***	
	Wald test: coefficient on (C) = (E)?				7.03***

Standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.10. Test statistic for Wald test is F test in Columns 1–3 and chi-squared test in Column 4. The sample for all models includes newly rated and never-rated firms.

[a] Compliance concerns refers to *hazardous waste* or *regulatory problems* KLD concerns.

[b] Emissions concerns refers to *substantial emissions* or *climate change* KLD concerns.

[c] Other concerns refers to agricultural chemical, ozone-depleting substance, or other KLD concerns.

[d] This negative binomial model includes firm-level conditional fixed effects.

DOI: 10.1002/smj

EXHIBIT A2: DESCRIPTION OF KLD ENVIRONMENTAL RATINGS (AS OF 2006)

KLD environmental strengths

1. **Beneficial products and services.** The company derives substantial revenues from innovative remediation products, environmental services, or products that promote the efficient use of energy, or it has developed innovative products with environmental benefits. (The term 'environmental service' does not include services with questionable environmental effects such as landfills, incinerators, waste-to-energy plants, and deep injection wells.)
2. **Pollution prevention.** The company has notably strong pollution prevention programs including both emissions reductions and toxic-use reduction programs.
3. **Recycling.** The company either is a substantial user of recycled materials as raw materials in its manufacturing processes, or a major factor in the recycling industry.
4. **Clean energy.** (previously called alternative fuels). The company has taken significant measures to reduce its impact on climate change and air pollution through use of renewable energy and clean fuels or through energy efficiency. The company has demonstrated a commitment to promoting climate-friendly policies and practices outside its own operations.
5. **Communications.** The company is a signatory to the CERES Principles, publishes a notably substantive environmental report, or has notably effective internal communications systems in place for environmental best practices. KLD began assigning strengths for this issue in 1996.[a]
6. **Property, plant, and equipment.** The company maintains its property, plant, and equipment with above-average environmental performance for its industry. KLD has not assigned strengths for this issue since 1995.
7. **Other strength.** The company has demonstrated a superior commitment to management systems, voluntary programs, or other environmentally proactive activities.

KLD environmental concerns

1. **Hazardous waste.** The company's liabilities for hazardous waste sites exceed $50 million, or the company has recently paid substantial fines or civil penalties for waste management violations.
2. **Regulatory problems.** The company has recently paid substantial fines or civil penalties for violations of air, water, or other environmental regulations, or it has a pattern of regulatory controversies under the Clean Air Act, Clean Water Act, or other major environmental regulations.
3. **Ozone-depleting chemicals.** The company is among the top manufacturers of ozone-depleting chemicals such as HCFCs, methyl chloroform, methylene chloride, or bromines.
4. **Substantial emissions.** The company's legal emissions of toxic chemicals (as defined by and reported to the EPA) from individual plants into the air and water are among the highest of the companies followed by KLD.
5. **Agricultural chemicals.** The company is a substantial producer of agricultural chemicals (i.e., pesticides or chemical fertilizers).
6. **Climate change.** The company derives substantial revenues from the sale of coal or oil and its derivative fuel products, or the company derives substantial revenues indirectly from the combustion of coal or oil and its derivative fuel products. Such companies include electric utilities, transportation companies with fleets of vehicles, auto and truck manufacturers, and other transportation equipment companies.
7. **Other concern.** The company has been involved in an environmental controversy that is not covered by other KLD ratings.

Source: KLD Ratings Methodology: http://www.kld.com/research/data/KLD_Ratings_Methodology.pdf.

[a] In 2005, after the period analyzed in this article, this issue was incorporated into the Corporate Governance Transparency rating.

Table A2. Performance improved most among firms with initial rating of poor: robustness tests

	Dependent variable:	1 Log emissions	2 Number of regulatory penalties
	Sample:	Newly rated and *always*-rated firms	Newly rated and *never*-rated firms
		OLS coefficients	Incident rate ratios
(A)	KLD rated × initial rating poor	−0.934***	0.564***
		[0.272]	[0.122]
(B)	KLD rated × initial rating mixed or good	0.117	1.217
		[0.133]	[0.214]
	Log employees	−1.322**	1.541***
		[0.607]	[0.254]
	Log sales	1.251***	1.006
		[0.463]	[0.178]
	Log assets	0.526	0.725
		[0.665]	[0.150]
	Log number of TRI-reporting facilities	1.063***	0.934
		[0.275]	[0.079]
	Year dummies (2000–2004)	Included	Included
	Firm-level fixed effects	Included	Included[a]
	Interactions between each control variable and the *ex ante* concerns comparison group dummy[b]	Included	
	Observations (firm-years)	3150	1089
	Firms	663	221
	R-squared (within)	0.17	
	Wald test: coefficient on (A) = (B)?	12.04***	11.29**

Column 1 displays OLS regression coefficients. Column 2 displays incident rate ratios from conditional fixed effects negative binomial regression models. Standard errors in brackets. *** $p<0.01$, ** $p<0.05$, * $p<0.10$. Test statistic for Wald test is F test in Column 1 and chi-squared test in Column 2; for both, null hypothesis is that the coefficients are statistically indistinguishable. These models are robustness tests for the results presented in Table 2.

[a] This negative binomial model includes firm-level conditional fixed effects.

[b] The *ex ante* concerns comparison group includes (a) always-rated firms, the KLD ratings of which in 1999–2000 included only KLD concerns (no KLD strengths), and (b) newly rated firms, the initial KLD ratings of which included only KLD concerns.

DOI: 10.1002/smj

Table A3. Rating effects moderated by regulatory stringency: robustness tests

	Dependent variable:	1 Log emissions	2 Number of regulatory penalties
	Sample:	Newly rated and *always*-rated firms	Newly rated and *never*-rated firms
		OLS Coefficients	Incident rate ratios
(A)	Highly environmentally regulated × KLD rated × initial rating poor	−1.072***	0.593*
		[0.343]	[0.184]
(B)	Highly environmentally regulated × KLD rated × initial rating mixed or good	0.783***	1.426
		[0.255]	[0.386]
	Highly environmentally regulated × log employees	−1.565**	1.584**
		[0.680]	[0.369]
	Highly environmentally regulated × log sales	1.096**	0.852
		[0.556]	[0.216]
	Highly environmentally regulated × log assets	0.972	0.888
		[0.815]	[0.262]
	Highly environmentally regulated × log number of TRI-reporting facilities	0.780*	1.075
		[0.401]	[0.166]
(C)	Low environmentally regulated × KLD rated × initial rating poor	−0.728	0.537**
		[0.479]	[0.163]
(D)	Low environmentally regulated × KLD rated × initial rating mixed or good	−0.147	1.092
		[0.159]	[0.258]
	Low environmentally regulated × log employees	−1.054	1.207
		[1.465]	[0.295]
	Low environmentally regulated × log sales	1.260	1.237
		[1.019]	[0.351]
	Low environmentally regulated × log assets	−0.048	0.703
		[1.236]	[0.197]
	Low environmentally regulated × log number of TRI-reporting facilities	1.320***	0.892
		[0.389]	[0.093]
	Year dummies (2000–2004) interacted with highly/low environmentally regulated	Included	Included
	Firm-level fixed effects	Included	Included[a]
	Interactions between each control variable and the *ex ante* concerns comparison group dummy[b]	Included	
	Observations (firm-years)	3139	1089
	Firms	660	221
	R-squared (within)	0.18	
	Wald test: coefficient on (A) = (B)?	18.89***	7.06***
	Wald test: coefficient on (C) = (D)?	1.33	4.81**
	Wald test: coefficient on (A) = (C)?	0.34	0.05
	Wald test: coefficients on (A) − (B) = (C) − (D)?	3.72*	0.13

Column 1 displays OLS regression coefficients. Column 2 displays incident rate ratios from conditional fixed effects negative binomial regression models. Standard errors in brackets. *** $p<0.01$, ** $p<0.05$, * $p<0.10$. Test statistic for Wald test is F test in Column 1 and chi-squared test in Column 2; for both, null hypothesis is that the coefficients are statistically indistinguishable. These models are robustness tests for the results presented in Table 3.

[a] This negative binomial model includes firm-level conditional fixed effects.

[b] The *ex ante* concerns comparison group includes (a) always-rated firms, the KLD ratings of which in 1999–2000 included only KLD concerns (no KLD strengths), and (b) newly rated firms, the initial KLD ratings of which included only KLD concerns.

DOI: 10.1002/smj

Table A4. Rating effects moderated by environmental efficiency: robustness tests

	Dependent variable:	1	2
	Sample:	Log emissions	Number of regulatory penalties
		Newly rated and *always*-rated firms	Newly rated and *never*-rated firms
		OLS Coefficients	Incident rate ratios
(A)	Less environmentally efficient × KLD rated × initial rating poor	−1.183***	0.530**
		[0.306]	[0.135]
(B)	Less environmentally efficient × KLD rated × initial rating mixed or good	−0.062	1.019
		[0.182]	[0.267]
	Less environmentally efficient × log employees	−1.231*	1.944**
		[0.634]	[0.577]
	Less environmentally efficient × log sales	1.271***	0.771
		[0.475]	[0.203]
	Less environmentally efficient × log assets	0.278	0.923
		[0.712]	[0.276]
	Less environmentally efficient × log number of TRI-reporting facilities	1.194***	0.866
		[0.295]	[0.132]
(C)	More environmentally efficient × KLD rated × initial rating poor	−0.094	0.64
		[0.577]	[0.304]
(D)	More environmentally efficient × KLD rated × initial rating mixed or good	0.307	1.061
		[0.199]	[0.275]
	More environmentally efficient × log employees	−2.095	1.624**
		[1.790]	[0.367]
	More environmentally efficient × log sales	1.191	1.03
		[1.454]	[0.258]
	More environmentally efficient × log assets	1.842	0.695
		[1.619]	[0.202]
	More environmentally efficient × log number of TRI-reporting facilities	0.551	0.838
		[0.670]	[0.113]
	Year dummies (2000–2004) interacted with less/more environmentally efficient status	Included	Included
	Firm-level fixed effects	Included	Included[a]
	Interactions between each control variable and the *ex ante* concerns comparison group dummy[b]	Included	
	Observations (firm-years)	2952	1026
	Firms	571	198
	R-squared (within)	0.178	
	Wald test: coefficient on (A) = (B)?	7.511***	5.13**
	Wald test: coefficient on (C) = (D)?	1.194	1.06
	Wald test: coefficient on (A) = (C)?	2.783*	0.12
	Wald test: coefficients on (A) − (B) = (C) − (D)?	1.041	0.07

Column 1 displays OLS regression coefficients. Column 2 displays incident rate ratios (IRR) from a conditional fixed effects negative binomial regression model. Standard errors in brackets. *** $p<0.01$, ** $p<0.05$, * $p<0.10$. Test statistic for Wald test is F test in Column 1 and chi-squared test in Column 2; for both, null hypothesis is that the coefficients are statistically indistinguishable. These models are robustness tests for the results presented in Table 4.

[a] This negative binomial model includes firm-level conditional fixed effects.

[b] The *ex ante* concerns comparison group includes (a) always-rated firms, the KLD ratings of which in 1999–2000 included only KLD concerns (no KLD strengths), and (b) newly rated firms, the initial KLD ratings of which included only KLD concerns.

DOI: 10.1002/smj