The concept of environmental performance and its measurement in empirical studies

Wolfgang Schultze · Ramona Trommer

Abstract We discuss the measurement of environmental performance (EP) in quantitative empirical research. Initially, we review and classify existing EP measures. Based on that, we analyze their validity and reliability. To provide a clear conceptualization of EP, we mainly refer to the framework of Wood (1991) and conceive EP as a multidimensional construct representing the extent to which companies meet the environmental expectations of their stakeholders. Finally, we discuss the operationalization of EP by examining stakeholders' expectations in detail and investigating qualitative characteristics of EP measures used within empirical research. Our analysis leads to the conclusion that measures based on inputs and outputs, operational processes and strategic EP provide construct validity. Generic EP measures used in large-scale studies should adequately represent stakeholders' environmental expectations, in particular referring to prospective indicators. Our study contributes to the research on EP measurement by providing an extensive literature overview, improving the theoretical understanding of the EP construct and providing basic recommendations for coherent EP measurement for empirical analysis.

Keywords Environmental performance measurement · Operational and strategic indicators · Stakeholder expectations · Multidimensional construct · Environmental impacts · Forward-looking indicators

1 Introduction

There is extensive empirical research on the relationship between environmental performance (EP) and corporate performance (CP). Understanding this relationship is

W. Schultze · R. Trommer (✉)
Department of Accounting, University of Augsburg, 86135 Augsburg, Germany
e-mail: ramona.trommer@wiwi.uni-augsburg.de
url: http://www.wiwi.uni-augsburg.de/bwl/schultze/

important, as companies are increasingly required to be both profitable and environmentally responsible (Lankoski 2000; Günther et al. 2004; Orlitzky et al. 2011). However, the results of these studies do not provide a consistent and comprehensive picture (for an overview, see Günther et al. 2004; Molina-Azorín et al. 2009). In particular, it is still not clear which factors mediate or influence the EP-CP relationship (Lankoski 2008). Deficiencies are mainly ascribed to methodological shortcomings of the studies and to problems regarding EP measurement (Salzmann et al. 2005).

Within empirical research a multitude of different EP indicators is used (Wagner 2001; Orlitzky et al. 2003; Wood 2010). Their choice seems to be mainly driven by practical feasibility. Only a few studies justify their measurement by theoretical considerations, e.g., by conceptualizing EP as companies' environmental outcomes (e.g., Lankoski 2000). However, there are numerous other EP measures not related to outcomes. Today it is not fully understood how these measures are causally connected, in order to explain the different results of the studies. Neither is it clear how they relate to the EP construct, that is, whether they measure EP at all. Many scholars criticize the lack of theoretical foundation (Salzmann et al. 2005; McWilliams et al. 2006). On a normative level the question arises how to measure EP correctly and several scholars call for further research into EP measurement (Günther et al. 2004; Weber 2008; Orlitzky et al. 2011). Wood (2010) recommends measuring EP more consistently before further conducting empirical studies, starting with a more thorough derivation of EP measures from theory.

We start our investigation with a review of currently used EP measures, considering the context in which they are used, and classify them into distinct groups. Regarding the quality of those measures, we derive two essential criteria from measurement theory, validity and reliability. A measure is valid when it measures what is intended to be measured and reliable when it provides an accurate and stable picture (Mitnick 2000; Kromrey 2009). A review of criticism raised against the different measurement categories shows major problems regarding the validity and reliability of EP measures. In order to obtain a more profound understanding of their relation to the (hypothetical) EP construct and to derive "good" measures, we reconsider the theoretical foundations of EP.

We conduct our analysis in three steps. First, we discuss the definition of the EP construct. A clear conceptualization does not exist so far, which can be considered a major reason for its heterogeneous measurement. The fundamental question in this context is: What exactly is the *performance* we want to measure? As performance is always related to the fulfillment of expectations, EP can basically be conceived as the extent to which companies meet the expectations of their stakeholders regarding environmental responsibility (Ruf et al. 1998; Carroll 2000). We specify this definition by applying the model of Wood (1991), which is one of the most influential and most comprehensive theoretical frameworks in this context (Orlitzky et al. 2003). Accordingly, EP is considered a multidimensional construct that does not only include environmental outcomes and impacts on the company, its stakeholders and the environment but also principles of environmental responsibility and processes of environmental responsiveness which determine future outcomes and impacts. However, we argue that the *impacts* of companies' behavior do not correspond to the basic performance definition. Companies are only expected to do what is within their power.

They are not held responsible for unforeseeable interferences outside of their control. We conclude that EP only refers to corporate behavior. Its external consequences, e.g., stakeholders' reactions to a company's emission reduction, represent different constructs. They are causally related to EP but underlie several influence factors beyond EP that, if not statistically controlled for, lead to biased empirical results.

In a second step of our analysis, we discuss how to operationalize EP based on this definition and distinction. To be able to measure the fulfillment of stakeholders' environmental expectations, these expectations should be known in detail. However, they are not substantiated by the model of Wood (1991). We therefore examine the expectations of key stakeholders. Besides a group called "environmental advocates" acting on moral beliefs, stakeholders intend to enforce their own interests, e.g., consumers expect to purchase healthy and safe products and investors demand the reduction of environmental risks in order to ensure the financial stability of a firm. The results indicate that environmental expectations are heterogeneous between different groups. The validity of EP measures depends on their adaption to three points: special interests of the stakeholder groups under examination (including their time horizon), special characteristics of companies/products and external factors relevant to stakeholders when forming their expectations. If large-scale studies neglect these points, e.g., by measuring "overall" stakeholder expectations or by not taking into account company characteristics properly, their EP measures are less valid. Nevertheless, in large-scale studies measures of EP applicable to numerous companies and representing the interests of multiple stakeholders are needed for reasons of practicability.

As a final step of our study, we conduct a detailed analysis of validity and reliability of different categories of EP with a special focus on the requirements of quantitative empirical research. By comparing the multidimensional EP model to the categorization of currently used EP measures, we identify five categories directly corresponding to the EP construct (i.e., providing construct validity): operational input indicators, output indicators, process indicators, indicators of strategic environmental management and indicators of environmental attitudes and objectives. As operational input and output indicators are closely related to the incurrence of environmental impacts, they are assumed to be most valid to EP measurement. However, our analysis reveals them to suffer from a number of shortcomings that in prior literature have not been taken sufficiently into account. We find stakeholders to be not only interested in a company's current but also in its *future* environmental impacts. Input and output indicators are retrospective and thus are not very useful in predicting future outcomes. In this respect, operational process indicators and strategic EP yield better results. Furthermore, single input and output indicators cover only small aspects of EP. Their aggregation to overall measures is highly subjective as their importance in causing environmental impacts (determining their weight) is still in dispute. They also mostly do not cover *all* environmental interactions of a company but only a representative selection. This must be seen critically, in particular with regard to a possible greenwashing behavior of companies, that is, the selective provision of beneficial information. While operational input and output indicators can be misleading, operational process indicators and strategic EP indicators usually provide a broad view of a company's EP which, if indicators can be classified as reliable, helps to differentiate "good" from "bad" performers. Strategic measures also score with high

comparability even between companies of different industries. In contrast, input and output indicators are very specific and their levels can only be compared within small peer-groups, e.g., sub-industries using the same technologies.

This allows drawing conclusions for EP measurement. First, EP measurement should only rely on the five categories of measures named above. Second, developing generic EP measures for large-scale studies comes with a loss in validity. Generic measures should at least correctly depict stakeholders' environmental expectations according to their relevance. In particular, as stakeholders are also interested in a company's future performance, prospective EP measures (e.g. indicators of strategic EP) should be represented adequately. Third, the consistency of empirical results can be further enhanced by better adjusting the choice of indicators to the special interests of the stakeholder groups involved, company/product characteristics and external factors influencing stakeholders' expectations.

Our paper contributes to the literature on EP measurement in two ways. First, our paper helps to improve the conceptualization of EP by pointing out the multidimensional character of EP and distinguishing it from related constructs. Second, our paper provides insights into valid and reliable EP measurement within empirical research by reviewing and classifying existing EP measures (in connection with the context in which they are used) and relating them to the dimensions of the EP construct. We do this by transferring Wood's model, which was originally designed for the measurement of corporate social performance (CSP), to EP measurement. Our results provide basic guidance for EP measurement.

The paper is structured as follows: The next chapter reviews the related literature and specifies measurement problems. Section 2.1 briefly describes the empirical research on the relationship of EP and CP and Sect. 2.2 classifies the EP measures used. Section 2 concludes with a critical evaluation of EP measures. The preceding chapter focuses on the definition of EP and discusses basic concepts. Furthermore, EP is distinguished from related constructs. Section 4 addresses the operationalization of EP in empirical studies. Based on a discussion of stakeholders' environmental expectations, quality characteristics of EP measures are analyzed. The paper concludes with final remarks.

2 The measurement problem

2.1 Related empirical research

A large number of studies examine the economic impacts of EP. Quantitative methods can be categorized in regression analyses, event studies and portfolio analyses (Günther et al. 2004; Salzmann et al. 2005; Ambec and Lanoie 2007).[1] Regression analyses, representing the vast majority of investigations on this topic, use multivariate data analysis to observe the impact of EP on CP. Early studies analyze the direct

[1] The studies discussed in the following were selected by scanning literature reviews (e.g., King and Lenox 2001; Wagner 2001; Ambec and Lanoie 2007; Dick et al. 2008; Dixon-Fowler et al. 2009) with the aim of providing a representative sample for each category of EP measures. We focused on studies of the last 20 years with a high impact on subsequent literature.

interrelation of EP and CP measures by asking "Does it pay to be green?". The intention behind the calculation of simple correlation coefficients is to uncover general principles which can be used to support future investment decisions (Schaltegger and Müller 1997). As empirical results show, the relationship between EP and CP is not as clear-cut as researchers expected. Even if the results often uncover positive economic impacts of EP (Hart and Ahuja 1996; Bhat 1999; Konar and Cohen 2001; Al-Tuwaijri et al. 2004), some studies also find non-significant (McWilliams and Siegel 2000) and negative relationships (Nehrt 1996; Cordeiro and Sarkis 1997; Halme and Niskanen 2001; Wagner et al. 2002). Though the results of these studies might be useful for capturing tendencies and trends, simple correlation studies are not able to provide explanations of how EP and CP are connected (King and Lenox 2001; Telle 2006). Furthermore they are unable to provide evidence on the question whether correlations are due to impacts of EP on CP. In fact, arguments for a reversed causality can also be made (Telle 2006). In addition, major factors of influence are not addressed (Schaltegger and Synnestvedt 2002).

To overcome these limitations, the research question was rephrased to "When and why does it pay to be green?". Corresponding studies focus on the causal relationship between EP and CP and its influencing factors. For example, the relationship of EP and CP is found to be mediated by innovativeness and environmental competitive advantages (Karagozoglu and Lindell 2000) as well as customer satisfaction (Luo and Bhattacharya 2006). Furthermore, complementary assets (Christmann 2000), management capability (Clarkson et al. 2011), innovativeness capability (Luo and Bhattacharya 2006), industry membership (Brammer and Pavelin 2006; Telle 2006), industry growth (Russo and Fouts 1997) and regulatory supportiveness (Karagozoglu and Lindell 2000) were identified to moderate the relationship. However, the causal links between EP and CP and their influencing factors are still not identified sufficiently (Schaltegger and Synnestvedt 2002; Marom 2006; Lankoski 2008).

As a second quantitative method, event studies assess the impact of positive and negative environmental incidents on abnormal stock returns (Salzmann et al. 2005). The results draw a mixed picture. Halme and Niskanen (2001) prove environmental investments to be negatively valued by investors, at least in the short-term. Contrarily, firms announcing to adopt the CERES principles experience positive abnormal returns (White 1996). In case of "bad news", e.g., environmental accidents or crisis, market reactions are consistently negative (Blacconiere and Patten 1994; Klassen and McLaughlin 1996). A high level of environmental disclosure, however, weakens the impact (Blacconiere and Patten 1994). The market also punishes high polluters (Hamilton 1995). Consistently, "good news", e.g., environmental awards, are rewarded with positive abnormal returns. Their occurrence is influenced by industry and time aspects (Klassen and McLaughlin 1996).

In portfolio analyses, the performance of environmentally responsible funds is compared to a benchmark (Pava and Krausz 1996; Salzmann et al. 2005). White (1996) shows that a portfolio of firms with above average environmental reputation earns higher risk-adjusted returns than the market in general. Other studies cannot find significant impacts (Hamilton et al. 1993; Statman 2000; Bauer et al. 2005).

In summarizing, we can conclude that empirical research is still far from providing a conclusive picture of why and how EP and CP are connected. An overview of the

results and methodologies of the discussed empirical studies is given in the Appendix. In the following, the measurement of EP within these empirical studies is examined.

2.2 Environmental performance measures in empirical research

Empirical research uses a large variety of EP measures. James (1994) groups them into process, resource consumption, emissions and waste, efficiency, risk, ecological impact, consumer perception and financial impact related measures. Bartolomeo (1995) develops a more systematic approach by separating performance indicators from impact indicators. Performance indicators are based on processes (efficient use of raw materials) and systems (effectiveness in achieving eco-efficiency targets) or refer to a financial dimension (economic efficiency in implementing environmental programs). Impact indicators evaluate environmental impacts in physical and monetary terms. Ilinitch et al. (1998) differentiate between internal systems measures (EP enhancing organizational processes), external stakeholder relations measures (interactions between the company and its external constituencies), external impact measures (negative environmental externalities), and finally, internal compliance measures (compliance with laws and regulation).

Our categorization builds upon Jung et al. (2001) and Günther et al. (2004). Basically, both separate the measures into a strategic and an operational category. The first category addresses a firm's attitudes and objectives regarding environmental responsibility as well as environmental management structures and processes, e.g. environmental information systems and environmental management systems (EMS). Jung et al. (2001) further separate operational measures into four causally linked subcategories, namely input, process, output, and outcome measures. Input and output measures refer to companies' physical interactions with the natural environment, e.g. raw material, water and energy consumption, land use, emissions, waste arising, and noise. Also the physical impacts of companies' products and services are considered. Process measures deal with process changes regarding the manufacturing process, the use of new technology, recycling activities, transportation, and supplier management. Also changes of inputs, package and design of products and services are included in these measures. Crucial is the category of outcome measures. Corresponding metrics do not measure inputs and outputs but their impacts on stakeholders, e.g., customers, employees, investors and society as a whole. This also includes economic impacts.

As the approach of Jung et al. (2001) does not cover all EP measures found in empirical studies, we amend it by further categories called "combined measures". These are self-calculated EP scores, perceived performance, environmental ratings, environmental funds, and environmental related events (similar see Günther et al. 2004[2]). They include measures that potentially span more than one operational or

[2]To cover measures calculated by the researchers themselves by gathering different indicators, we included a new category "self-calculated scores". Günther et al. (2004) also suggest measures referring to the level of environmental reporting as a separate category. The level of environmental reporting itself and partly also its correlation with EP and CP are the subject of empirical research (e.g., Al-Tuwaijri et al. 2004; Clarkson et al. 2008). However, its use to measure EP is substantially doubted as it rather provides a self-constructed image (McGuire et al. 1988). Recent studies rarely make use of EP measures based on the level of environmental reporting. Therefore, we removed this category.

Single indicators				
Operational	**Input oriented** • Raw material, water and energy consumption • Land use	**Process oriented** • Process changes (e.g., manufacturing process, use of new technologies, recycling, supplier management, transportation) • Product changes (e.g., ingredients, packaging, design)	**Output oriented** • Emission of air, water, and land pollutants (e.g., toxic emissions, waste, radioactive substances, noise) • Environmental compatibility of products	**Outcome oriented** • Impacts on stakeholders • Impacts on the company (e.g., law suits, fines and penalties, environmental liabilities)
Strategic	• Environmental attitudes and objectives • Strategic environmental management			

Combined measures	
Self-calculated scores	• Weighted combination of indicators from different categories to a single measure
Perceived performance	• Assessment of overall environmental responsible behavior based on surveys
Environmental ratings	• Third party assessment of a companies' environmental responsible behavior based on qualitative and quantitative criteria
Environmental funds	• Third party assessment of a companies' environmental responsible behavior based on fund-specific criteria
Environmental related events	• External available information on environment-related incidents directly or indirectly affecting companies or whole industries

Fig. 1 Categorization of environmental performance measures

strategic category. Nevertheless, single measures can also concern a specific category, e.g., self-calculated scores can be based on output indicators or an environmental event could refer to the implementation of an EMS.

The resulting categorization provides a systematic overview of EP measures used in empirical research (see Fig. 1).

The category of operational EP measures comprises indicators of short term environmental improvement activities (Günther et al. 2004). Some studies use input measures that focus on the consumption of raw materials, water, energy, and land (e.g., Wagner 2005b). Other studies apply process measures referring to process changes that lead to lowered environmental impacts. King and Lenox (2002), for example, examine different methods of pollution prevention (waste prevention, treatment and transfer). Al-Tuwaijri et al. (2004) operationalize EP by the recycling quota of toxic waste. Some studies also apply financial indicators. E.g., Nehrt (1996) evaluates the use of new pollution control technology by financial investments.

Output measures are most commonly used in empirical studies. As companies' outputs affect the environment in various ways, this category comprises a large number of indicators. Measures are mostly quantified in physical units (e.g., carbon dioxide emissions in parts per million, physical waste arising in tons, fresh water consumption in cubic meter, land use in hectare, etc.), but are often scaled by

production volume. Data on the absolute or relative amount of toxic releases are mostly publicly available. This might explain the high frequency of data from the toxic-release-inventory (TRI) (e.g., Hart and Ahuja 1996; Cordeiro and Sarkis 1997; Griffin and Mahon 1997; Bhat 1999; King and Lenox 2001; Konar and Cohen 2001; King and Lenox 2002; Clarkson et al. 2011). Data of other national emission registers are used as well (e.g., Wagner 2005b; Telle 2006). Occasionally, data collection is based on surveys, e.g., to assess GHG emissions (Busch and Hoffmann 2011) and the percentage of chlorine-free pulp production (Nehrt 1996). As stated above, the environmental compatibility of products and services is also relevant to operational EP measurement. For example, Sammer and Wüstenhagen (2006) examine the eco-labeling of washing machines by their water and energy consumption. Some empirical studies operationalize EP by outcome measures referring to impacts of environmental protection activities. An example is the measurement of EP by environmental lawsuits (Konar and Cohen 2001). Environmental reputation also represents an outcome for the company. Accordingly, reputation rankings such as Fortune Magazine's "United States' most admired companies" (FAMA) (Griffin and Mahon 1997; Luo and Bhattacharya 2006) and Management Today's "Britain's most admired companies" (BMAC) (Brammer and Pavelin 2006) can be assigned to this category. White (1996) measures environmental reputation by ratings of the Council on Economic Priorities (CEP).

Within the category of strategic EP measures, indicators mostly refer to the adoption of environmental management systems (Dowell et al. 2000; Melnyk et al. 2003; Dick et al. 2008). Banerjee (2002) evaluates EP on the basis of environmental orientation and strategy. Busch and Hoffmann (2011) assess the existence of a carbon management system and carbon reduction targets and strategies.

As previously stated, some measures combine indicators of different categories to an overall score and thus cannot be assigned to a single category of operational or strategic EP measures. Their use is based on the conclusion that relying on single indicators does not cover the complex EP construct sufficiently. Hence, they calculate scores by combining and weighting different EP indicators. E.g., Christmann (2000) evaluates "best practices" of environmental management by the use of pollution-prevention technologies, the innovation of proprietary pollution-prevention technologies and an early timing of environmental strategies. Furthermore, Brammer and Pavelin (2006) use data on environmental policies, systems, reporting, and impacts to measure environmental related issues.

The category of perceived performance refers to the perception of environmental performance by business representatives (e.g., CEOs, business unit managers).[3] Data is mainly collected by company surveys. Surveys partly gather objective facts about operational and strategic EP indicators (see above), but also assess personal views, e.g., when asking to rate the level of EP relative to businesses' competitors (Günther et al. 2004). For example, Judge and Douglas (1998) measure EP by the perceived compliance with environmental regulations, environmental management and reporting strategies. Biondi et al. (2000) asked the participants of their survey for their environmental management tools and methods. Karagozoglu and Lindell (2000) assess

[3]Some other EP measures such as environmental reputation are perception-based by nature.

managers' perception of energy use, resource use and pollution prevention compared to their companies competitors. Finally, the questionnaire of Álvarez Gil et al. (2001) includes items on environmental costs and savings and the level of environmental training programs, green purchasing policies, communication policies, customer cooperation, energy and water-saving actions, and recycling activities.

The category of environmental ratings differs from the former categories of combined EP measures as these ratings are based on the assessment of third parties. Environmental ratings provided by rating agencies, for example the Franklin Research and Development Corporation (FRDC) (Russo and Fouts 1997), Kinder, Lydenberg and Domini (KLD) (Griffin and Mahon 1997; Waddock and Graves 1997) and the Swiss bank Sarasin & Cie (Ziegler et al. 2002), are often used to measure EP.

Environmental funds are mainly applied in portfolio analyses in order to separate "good" from "bad" performers. Similar to ratings, the assignment of companies to environmental funds represents a third party assessment, in this case to identify firms that comply with socially responsible investment (SRI) strategies. As portfolio analyses compare the performance of stock market portfolios, they basically revert to mutual funds using environmental criteria for the inclusion or exclusion of companies (Hamilton et al. 1993; Statman 2000; Bauer et al. 2005).

Environmental related events as used in event studies rely on certain incidents that become public. Depending on the context, event studies use information releases regarding toxic emissions (Hamilton 1995), environmental awards (Klassen and McLaughlin 1996), environmental accidents and crises (Blacconiere and Patten 1994; Klassen and McLaughlin 1996) and environmental investment announcements (Halme and Niskanen 2001) to observe the effects on daily abnormal returns.

The review shows that a large variety of EP measures is used within empirical research. To revert to the inconsistency of empirical results that was discussed in the last section, this can mainly be ascribed to methodological shortcomings: Either differences in the relationship between EP and CP are not taken fully into account or differences in the research process bias the results (Wagner 2001; Allouche and Laroche 2005). Particularly, problems regarding EP measurement play an important role (Carroll 2000; Wagner 2001; Salzmann et al. 2005). E.g., nature and strength of the relationship between EP and CP depend fundamentally on EP measurement (Griffin and Mahon 1997; Orlitzky et al. 2003; Günther et al. 2004; Busch and Hoffmann 2011). Contradictory results even appear in cases based on the same company sample (e.g., Nehrt 1996). Major shortcomings of EP measures are discussed in the following section.

2.3 Critical evaluation of environmental performance measures

The construct of EP has to be operationalized by observable indicators. Measurement theory provides two main criteria determining the quality of measures: validity and reliability (Mitnick 2000; Kromrey 2009). A measure is valid if it measures what it is supposed to measure ("measure the right things"). Therefore, it should be based on sound theoretical foundations (Carroll 2000). To validly operationalize the construct, indicators should cover it correctly and completely (Bartolomeo 1995; Kromrey 2009). To be reliable, a measure should deliver an accurate and stable picture of EP ("measure the things right"). This ensures data collection and analysis to be

reproducible by other researchers, with other methods and, under stable conditions, at a different point in time (Kromrey 2009). First of all, an accurate representation of the situation requires objective data (Tyteca 2002). To allow comparisons in time, among similar units or with a reference value, indicators should further be consistently measured and sufficiently quantified (Wagner 2005a). Reliable measurement also depends on data availability (Tyteca 2002). This includes the question if the required information can be observed with a reasonable effort or if the companies or external information sources (e.g., governmental agencies, research institutes) already possess the required data and if they are willing to cooperate.

As environmental disclosure is mostly voluntary and not standardized, the availability of objective and comparable data is low (Christmann 2000; Berkhout et al. 2001). Organizations like the Global Reporting Initiative (GRI) and the International Standard Organization (ISO) provide measurement guidelines but their adoption is voluntary (Jung et al. 2001). Furthermore, an obligation to fully correspond to their requirements or to explain deviations in detail does not exist. Therefore, the reliability of the disclosed information is questionable (Brewer and Stern 2005; Luo and Bhattacharya 2006; Promberger and Spiess 2006). Company surveys are often considered an alternative. Even if they are able to capture facts that would hardly be available otherwise and score with a high flexibility according to their scope and sample, information on perceived performance also lacks objectivity as it mostly cannot be verified by the researcher. Furthermore, as questionnaires are tailored to the special scope of each study, reproducibility and comparability of the measures are low (Günther et al. 2004).

Also mandatorily provided information is subject to criticism. In particular, TRI data is criticized for having undergone many changes during the years and thus being hardly comparable. Furthermore, it is provided by companies themselves and may not be measured reliably (Wood 2010). Mixing TRI data with data from other emission registers (e.g., Wagner 2005b) aggravates this problem. TRI data also disregards non-toxic energy and material flows that might affect the environment negatively, too (Ambec and Lanoie 2007). As the high number of studies using TRI data shows, the operationalization of EP is rather driven by data availability than by selecting well-founded measures. EP measurement seems to follow the principle "take what you can get".

Beyond TRI data, previous studies use a multitude of measures based on various, substantially limited data sources. Lankoski (2000) notes difficulties when EP is measured by environmental investment expenditures. Particularly, process investments involve limited observable costs such as overhead costs and costs resulting from quality impacts or regulatory delays. But even if costs could be completely captured, the relevance of process indicators is generally questionable as they are not directly related to outcomes. The spending might be inefficient and not reduce harmful environmental impacts (Wagner 2005b; Busch and Hoffmann 2011).

Third party assessments mostly do not provide a better measurement basis. As evaluation criteria and weightings of reputation indices and environmental ratings are often not public, they are criticized as being subjective and hard to evaluate (Ruf et al. 1998; Lankoski 2000; Ziegler et al. 2002). Many researchers find that reputation indices based on the perceptions of business executives and analysts represent

a companies' reputation in the business community rather than to provide a valid measure of the assessment of all stakeholders (Wood and Jones 1995; Mitnick 2000; Peloza and Papania 2008). Also certifications of environmental managements systems (EMS) according to ISO 14001 or EMAS are not regarded as a sufficient EP measure. Even if third party audits based on publicly available guidelines guarantee some degree of objectivity, they do not specify the improvement exactly (Christmann and Taylor 2001). Data availability might be good (Wood 2010) but EMS guidelines are criticized for their "openness" (Schaltegger and Synnestvedt 2002). Finally, the assignment to ethical and ecological funds underlies fund-specific criteria, for instance the formulation of an environmental plan or the exclusion of companies involved in nuclear power (Günther et al. 2004). These assessments are partly very rough and criteria are often not hard to fulfill (Wood and Jones 1995). Furthermore, fund affiliation provides binary information, i.e., it is not possible to rank companies according to their EP level. Thus, environmental funds do mostly not deliver sophisticated information on the level of EP.

Notably some measures are used ambiguously (Wood 2010). E.g., while Luo and Bhattacharya (2006) use the FAMA reputation ranking to measure social performance, others treat reputation as an outcome of social performance (Brammer and Pavelin 2006). A similar example is environmental innovativeness. While Adamczyk et al. (2009) explicitly suggest to evaluate EP on this basis, Karagozoglu and Lindell (2000) see environmental innovativeness as a result of EP.

The extent to which single measures cover the EP construct differs remarkably. Self-calculated scores or third-party assessments mostly include multiple aspects of EP and thus provide a relatively broad assessment. In contrast, many proxy measures focus on narrow aspects of EP (Günther et al. 2004; Salzmann et al. 2005; Busch and Hoffmann 2011). Their use is justified by the (questionable) assumption that they are representative for the total EP (Lankoski 2000).

These examples show major problems regarding the validity and reliability of EP measures. The heterogeneity of the EP measures is often ascribed to a lack of theoretical foundation (Wood and Jones 1995). They do not automatically capture the underlying construct to the same extent (Carroll 2000) and their conceptual connections are not yet fully understood (Ilinitch et al. 1998; Wagner 2005a). As a growing body of literature deals with the economic impacts of EP, further research on EP measurement is strongly needed (Wood 2010). But how can EP measurement be improved? According to the criterion of validity, indicators should cover the EP construct completely and correctly. A clear and unambiguous definition of EP forms the basis of a sound measurement. Measurement problems can finally be ascribed to the vague definition of EP (Salzmann et al. 2005)[4]. McWilliams et al. (2006) conclude that "it is impossible to measure what we cannot define and, as long as we use different definitions, we will get empirical results that cannot reliably be compared" (p. 10). Hence, before turning to measurement issues, we need to address the definition of EP.

[4]For an overview of different definitions of EP, see Günther et al. (2004).

3 Reconsidering the environmental performance construct

3.1 Multiple dimensions of environmental performance

Literature does not provide a clear understanding of EP. When measuring EP, many empirical studies do not refer to a definition at all (Ilinitch et al. 1998). A universal understanding should start from the term *performance* which is defined as the extent to which companies achieve their principals' targets, i.e., fulfill their expectations (Milgrom and Roberts 1992). The perception of EP as the fulfillment of environment-related requirements of stakeholders can therefore be regarded as a fundamental definition of the construct (similar see Wagner 2005a). But who are the environmental stakeholders of a company? And what do they expect? According to Freeman (1984), a company should satisfy the interests of all groups that are influenced by it or that can influence its actions. Environmental responsibility follows this multi-stakeholder perspective (Wood 1991; Atkinson et al. 1997; Peloza and Papania 2008). Judge and Douglas' (1998) definition derives from these societal obligations and describes EP as "a firm's effectiveness in meeting and exceeding society's expectations with respect to concerns for the natural environment." And as companies are responsible not just to legislative bodies but to various environmental stakeholders, EP extends "beyond mere compliance with existing regulations to a proactive stance concerning future environmental considerations" (p. 245).

To specify stakeholder expectations, some definitions of EP (even if most of them do not explicitly refer to this issue) directly focus on an outcomes perspective. For example, Lankoski (2000) states that "the concept of environmental performance pertains to the level of harmful environmental impact caused by the activities a firm" (p. 10) and that "environmental performance is a vector of all [environmental] impacts. Environmental impacts occur through land use, resource use, and pollutant releases into air, water, and land throughout the life-cycle of a product" (p. 11).

Other researchers understand EP as a concept that comprises further dimensions beyond outcomes. In this context, the most cited concept is Wood's (1991) CSP model.[5] As EP can be seen as a subcategory of CSP, this model can be directly applied to the EP construct (Ilinitch et al. 1998; Busch and Hoffmann 2011). Wood (1991) defines CSP as "a business organization's configuration of principles of social responsibility, processes of social responsiveness and policies, programs, and observable outcomes as they relate to the firm's societal relationships" (p. 693). These three dimensions build upon each other: structural principles of responsibility are the inputs, processes of social responsiveness are the throughputs and outcomes and impacts are the results of performance (Wood 2010). As Wood's concept considers the companies to be responsible to fulfill their stakeholders' needs, it is directly related to the general concept of performance discussed above.

Principles of environmental responsibility as the first dimension include legitimacy, public responsibility and managerial discretion. Legitimacy emphasizes the need to adjust business actions to the demands of stakeholders. The responsibility of

[5] Orlitzky et al. (2003) see Wood's (1991) definition as "one of the most influential, helpful, parsimonious, and yet comprehensive conceptualizations of CSP" (p. 411).

a firm refers to society and a company's stakeholders and is not only limited to "fixing what it has broken" but also extends to "avoiding future breakage, and [. . .] helping to solve those social problems that affect it" (Wood 1991, p. 699). When companies do not fulfill their obligations, society has sanction mechanisms. E.g., consumers might refuse to buy a company's products, employees might leave the firm and stockholders might sell the stocks. Wood's process dimension is related to the existence of structures of social (environmental) responsibility and processes as environmental scanning, stakeholder management and issues management. Finally, the outcomes and impacts dimension as the best observable part refers to the societal outcomes and impacts of corporate actions. As changes in physical inputs, outputs and operational processes are determined by the structural principles of social responsibility and produced, monitored and evaluated by the processes of corporate social responsiveness, they are perceived as the "outcomes" of these principles and processes.

EP definitions focusing on environmental impacts of a company only refer to the outcomes and impacts dimension and therefore can be classified as one-dimensional. Wood (2010) also states that, "in a sense, outcomes and impacts are what CSP is all about" (p. 69). Outcomes and impacts are the results of EP, i.e., what is expected to be. But does that mean principles and processes can be left out of consideration? Their existence is not bound to current operational results but rather guarantees flexibility and lowered environmental impacts in the future (Klassen and McLaughlin 1996; Lankoski 2000; Günther et al. 2004). Therefore, principles of environmental responsibility and processes of environmental responsiveness can be considered indicators for outcomes and impacts in the future. Leaving them out means abandoning important information on future performance (Mitnick 2000; Agle and Kelley 2001). Accordingly, Wood (2010) concludes that outcomes and impacts cover "only a very narrow slice of the larger, more comprehensive concept" (p. 54).

3.2 Distinction from related concepts

A clear distinction from related concepts enhances an unambiguous measurement. In this context we particularly refer to the system boundaries of EP. While the major part of performance in Wood's model, namely principles, processes and outcomes, refers to the company itself, impacts represent the consequences of this company-based performance for stakeholders, environment and the company. Are they really part of the EP construct? In other words, do stakeholders' expectations extend to the *impacts* of companies' behavior? According to its basic definition, performance, representing the degree to which companies fulfill their stakeholders' targets, refers to actions taking place at the company level (see Sect. 3.1). If a company has undertaken serious efforts to reach the targets of stakeholders but due to external influences that could not have been foreseen could not achieve the desired impacts, stakeholders usually will not punish it for this failure. Hence, we conclude that their expectations usually do *not* refer to issues outside the power of a company. Therefore, consequences of companies' environmental behavior as far as they can also be influenced by external factors do not represent EP but different (even though closely related) constructs. In the following we will illustrate how these conceptual differences affect EP measurement.

Wood and Jones (1995) identified potential connections between companies and stakeholders. Stakeholders do not only "establish expectations" but also "experience the effects of corporate behavior", "evaluate the effects or potential effects of corporate behaviors on their interest", and "act upon their expectations, experiences and evaluations" (p. 243). Wood and Jones assume all of these stakeholder roles to be relevant to measurement, as they are addressed by indicators used in empirical research (for an overview, see ibid.).[6] However, we argue that stakeholder expectations, perceptions, evaluations and actions are not equally qualified to measure EP. In particular, stakeholders' expectations and corresponding perceptions, evaluations and actions might be not fully determined by the EP level but influenced by other factors, too.

In particular, stakeholders' expectations set a benchmark to the environmental responsible behavior of a company, but their realization depends on various company-specific and contextual variables such as regulatory forces, growth prospects, the possibility to achieve competitive advantages (e.g., cost advantages), top-management commitment, asset availability, and financial resources (e.g., Banerjee et al. 2003; Clarkson et al. 2011). Also stakeholders' perceptions and their evaluations of how companies meet their expectations do not only depend on the actual level of performance. To some extent a companies' EP can be directly observed by interested stakeholders; sometimes intended by the company (e.g., tree planting), but sometimes also unintended (e.g., oil spills). But to a larger part stakeholders have to rely on secondary information. Information sources are inside (e.g., environmental reporting, eco-labeling) and outside the firm (e.g., media reports, third-party assessments such as eco-certifications) (Bloom et al. 2006). This information could be biased. Furthermore, the perception and evaluation process is based on the selection of information sources and is influenced by several factors, e.g. stakeholder characteristics (stakeholder types, issue support, social value orientation) and company characteristics (reputation, industry, marketing strategies) (Du et al. 2010). This also distorts the information processing. The same applies to stakeholders' actions. A firm which shows that its ecological footprint is low might create a higher willingness to pay for its products among customers, attract employees, lower its eco-related taxes, and attain positive media coverage (Lankoski 2000; Telle 2006; Ambec and Lanoie 2008; Weber 2008). However, stakeholders' decisions do not only depend on the perception and assessment of EP but are affected by other factors as well. Using the example of the customers' buying decision, also prior consumption experience and the assessment of innovativeness capability and product quality might influence the results (Luo and Bhattacharya 2006).

Besides direct cost effects of environmental protection, these interactions of companies and stakeholders represent the fundamental connection between EP and its impacts on CP (Wood 2010). Actions focusing on issues that do not correspond to shareholders' expectations, that are not perceived by stakeholders or that do not affect stakeholders' actions will not enhance CP (Peloza and Papania 2008). Figure 2

[6] Wood and Jones (1995) suppose that *different* stakeholder groups set the expectations, experience the effects and evaluate the outcomes. However, when a stakeholder bases his decisions upon the assessment of how well a company fulfilled his expectations, he has to perceive and evaluate this company's behavior on his own, too.

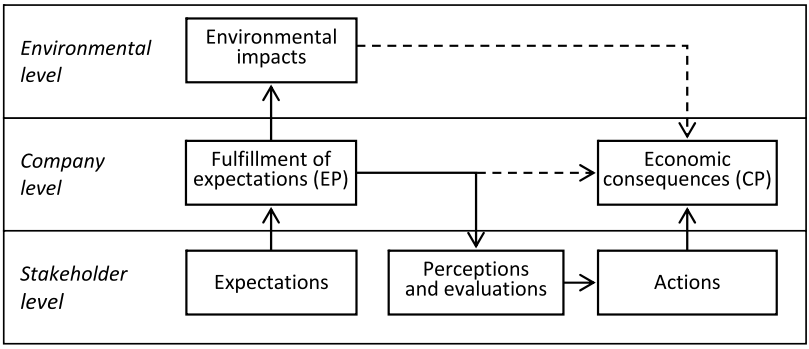

Fig. 2 Basic links from stakeholder expectations to CP

summarizes the described interactions of stakeholders, companies and the environment.

Output measures referring to the stakeholder level, namely stakeholder perceptions and evaluations of EP (e.g., environmental reputation) and their corresponding actions (e.g., legal actions, consumer choice) need to be distinguished from EP as they belong to different constructs, even though they are closely related.[7] As long as it is not possible to control for other factors affecting their level, they can only provide biased EP measures. When exploring the causal connection of EP and CP, they should rather be used as moderating variables and their influencing factors should be treated as mediating variables.

In summarizing this chapter, we discussed the definition of the EP construct and distinguished it from related concepts. We observed that EP can generally be defined as the extent to which companies meet the environmental expectations of their stakeholders. Accordingly, EP measurement needs to reflect the fulfillment of expectations of relevant stakeholders. Even if this insight is nothing new (e.g., see Ruf et al. 1998; Carroll 2000), it still has not reached measurement practice. Thus, Wood (2010) explicitly advices scholars to refocus on stakeholders when measuring CSP. However, none of the definitions discussed above gives advice how to operationalize EP as they do not tell exactly what stakeholders *expect*. Therefore, the next section starts with analyzing stakeholder expectations. This is followed by drawing specific conclusions for EP measurement within empirical research.

4 Operationalizing environmental performance

4.1 What do environmental stakeholders expect?

Considering that performance refers to the fulfillment of objectives, knowledge about the latter is crucial for the measurement of performance. To get to the bottom of EP

[7] These insights can be related to our previous observation that some EP measures, e.g., environmental reputation, are also used as dependent variables in other EP-CP studies (see Sect. 2.3).

measurement, the basic question is: "To whom are corporations responsible, and for what?" (Wood 2010, p. 69). The environmental expectations of several stakeholder groups might differ (Banerjee 2002; Wagner 2005a; McWilliams et al. 2006; Peloza and Papania 2008).

Initially, a group called "environmental advocates" (Walter 2005, p. 41) acts on moral beliefs without being directly affected by a company's environmental actions. Potentially all stakeholder groups (e.g., customers, employees, suppliers, business partners, investors) might request a reduction of externalities based on their moral beliefs (Sturm 2000). In particular, non-governmental organizations (NGOs) monitor environmental offenses in order to put pressure on those companies (Tyteca 2002). Furthermore, the government has a moral duty to act for "weak" stakeholders, e.g., future generations. Negative externalities also can lead to environmental damages that have to be fixed by spending public money. Thus, the government enforces environmental regulation (Walter 2005; Promberger and Spiess 2006).

Besides moral considerations, many stakeholders are interested in environmental responsible behavior for their own reasons. If companies' environmental behavior negatively affects residents' interests, e.g., healthiness and safety of their living environment, they might establish interest groups to enforce their claims (Tyteca 2002; Walter 2005). Employees are affected by a company's environmental behavior, too. They request prevention of accidents and pathogenic work conditions (Wood 2010). Customers are primarily interested in buying high-quality products at low prices (Promberger and Spiess 2006). Product quality involves environmental criteria (e.g., avoiding food and textiles containing pollutants). Furthermore, ecological characteristics occasionally determine follow-up costs of products (e.g., running and disposal costs).

Stakeholders keeping contractual relations to a company are interested in its financial stability. E.g., customers want to secure long-term product availability and after buy services, employees demand job security, suppliers and business partners want stable and secure business relationships and banks want companies to lower their financial risks (Wagner 1997; Tyteca 2002). As environmental risks threaten financial stability, all contracting partners demand their reduction. Economic stability is also relevant to the government as companies pay taxes and provide jobs (Sturm 2000; Promberger and Spiess 2006). Furthermore, customers and employees want to reach social recognition when doing business with a publicly appreciated company (Promberger and Spiess 2006). Also other contracting partners might expect to benefit from image spillovers (Walter 2005).

Owners are primarily interested in financial benefits, while avoiding financial risks and liabilities due to non-compliance with environmental regulations (Wagner 1997; Promberger and Spiess 2006). Furthermore, the profitability of their investments depends on good stakeholder relationships. This involves secondary objectives such as environmental responsible behavior helping to encourage stakeholders to achieve owners' primary objectives (Atkinson et al. 1997; Wagner 1997). However, as environmental responsibility is often very costly, owner's interests partly conflict with those of stakeholders. To their advantage they are able to evaluate their company's EP better than most stakeholders (Wood and Jones 1995). Stakeholders are often not able to distinguish good performers from poor performers that only mimic responsible behavior in order to write extra-profits (often referred to as "greenwashing") (Banerjee

2002; Clarkson et al. 2008). Thus, a further requirement of all environmental stakeholders refers to transparency and credibility of business actions (Reinhardt 1998; Miles and Covin 2000).

In summary, stakeholders demand a company to:

- reduce environmental externalities (environmental advocates)
- comply with regulations (government)
- avoid negative health and safety effects (neighbors, employees, consumers)
- reduce environment-related follow-up costs of products (consumers)
- reduce environmental risks (contracting partners, government)
- increase environmental reputation (contracting partners)
- increase transparency and credibility (all stakeholders)

These points represent basic categories which can be further subdivided. E.g., environmental advocates can explicitly refer to climate protection, wildlife conservation, the protection of nature reserves, etc. Again, a further differentiation within these segments into sub goals is possible.

The fulfillment of expectations that do not apply to the concerned stakeholders would not cause reactions. Conversely, not considering relevant expectations would disregard relevant information. Operationalizing EP by "wrong expectations" would therefore bias the correlation of EP and the dependent variable. Along with that, Wood and Jones (1995) find a serious mismatch of "variables which are mixed and correlated almost indiscriminately with a set of stakeholder-related performance variables that are not theoretically linked" (p. 231). They call this phenomenon "stakeholder mismatching". As an example, they discuss the impact of poor pollution performance on accounting measures of FP (indicating customer expectations) and stock prices (indicating owners' expectations). Owners might be concerned of negative effects on future cash flows, e.g., clean-up costs, technology upgrades, and regulative measures and therefore react negatively. Four marked-based studies supporting this assumption are cited. In contrast, studies examining the impact on accounting measures show no significant relationships. Wood and Jones explain this by pollution rankings usually not having a direct effect on purchase decisions of consumers.

The analysis conducted above can be used to examine customers' expectations further. Environmental expectations of consumers are primarily related to the product. They demand high quality and safety (e.g., organic food, bulbs not containing mercury) and low follow-up costs (e.g., low energy consumption of fridges, low CO_2 emissions of cars since they are taxed). Thus, eco-friendliness often corresponds to the own interest of consumers. Ethical issues also play an important role. E.g., even though it is more expensive, many consumers use eco-electricity. Others prefer local products in order to reduce their transport distances. Furthermore, many consumers are not only interested in the products/services but also in the companies providing them. Firms claiming to be environmentally responsible (e.g., by running rainforest campaigns or by releasing mission statements) can often increase their sales. In line with the recent sustainability debate, many consumers particularly attach high values to actions lowering environmental externalities or reducing risks over the long run, e.g., the changeover to regenerative energy sources. In contrast, conspicuously negative performance coming to light might lead to negative consumer reactions and even

consumer boycotts. This mainly refers to actions causing environmental damages, e.g. illegal waste disposal or abnormally high emissions. In assessing all these issues, consumers rely on company sources, e.g., advertising and product-specifications, but also on external sources such as test reports, eco labels, eco awards and media reports.

These examples reveal a number of points to consider in the measurement of EP when studying customer-related effects on FP. In-depth analyses of customer expectations include the following questions: What do customers precisely expect regarding the environmental friendliness of a specific product/service (e.g., regarding product composition, functional characteristics, recyclability) and how do they weight the different aspects? Are they willing to pay more for the reduction of environmental externalities beyond direct increases in product/service quality? What do they expect from the company providing the product/service, e.g., regarding its environmental impacts and its environmental attitudes, objectives, and policies? How important is future performance to them (e.g., ecological sustainability, ecological risks)? All these questions need to be answered in order to measure EP in a particular case.[8] In summary, it can be concluded that:

- each stakeholder group has specific environmental expectations differing at least partly from those of other groups,
- expectations can refer to both current and future performance,
- expectations are directed at the specific environmental interactions of products/services and companies depending on their special characteristics (e.g., composition and function of products; technologies and locational factors of companies); they might also depend on external factors (e.g., legal pollution limits, technological possibilities),
- as stakeholders' interests, company/product characteristics and external factors might change over time, stakeholder expectations are changeable, too,
- to evaluate EP, stakeholders are faced with extensive information needs which they possibly also cover by other than company sources; these external sources (e.g., media, interest groups) mediate the relationship of EP and stakeholder perceptions,
- stakeholders' evaluations are usually also affected by factors other than EP (in case of customers, e.g., product quality, general company image and financial stability) whose influence should be controlled in order to detect the true relationship of EP and FP.

Hence, a precise operationalization of EP would take into account all factors influencing the expectations of stakeholders in the particular case: the interests of the stakeholder groups under examination, specific product and company characteristics and relevant external factors. Furthermore, to measure the true relationship of EP and FP, researchers need to consider the special context of each study, including all factors moderating and mediating the investigated connections.

[8] An example for stakeholder expectations determining their willingness to pay for "green" products is provided by Azzone and Noci (1996).

4.2 Implications for EP measurement

In the following, we finally discuss the measurement of EP within empirical studies. A special difficulty of quantitative empirical research lies in incorporating large company samples with a heterogeneous stakeholder structure. Stakeholder structure varies remarkably between industries (Álvarez Gil et al. 2001; Roome and Wijen 2006), and can be highly variable even within a single industry (Peloza and Papania 2008). While companies are able to examine their stakeholders' expectations in detail and, based on that, assess their fulfillment, large-scale studies need to apply measures that abstract from the characteristics of single firms (Ruf et al. 1998; Wood 2010) without losing their comparability. According to the insights provided above, valid EP indicators require considering the scope of each study as far as possible. If researchers focus on single impact mechanisms involving certain stakeholder expectations (e.g., purchase or investment decisions) or certain company characteristics (e.g., special industries or countries), they should account for that by selecting appropriate EP measures. If empirical studies investigate multiple impact mechanisms involving different or ex ante unknown stakeholder expectations, however, measures need to reflect a common position of all stakeholders. Such a generic EP measurement concept ideally consists of broadly applicable metrics that focus on core areas of EP, i.e., information relevant to key environmental stakeholders (Wagner 2005a). Carroll (2000) sees the challenge in developing measures which "rely on stakeholders' opinions or assessments of performance and then try to configure them into some kind of overall measure that will provide a coherent depiction" (p. 473). In this context, measures assessing a company's overall environmental impacts are considered highly valid (Ruf et al. 1998; Lankoski 2000).

To correctly depict the EP, researchers also need to pay attention to the reliability of measurement.[9] Particularly, the combination of specific stakeholder expectations to overall measures (in order to be applicable to a large company sample) requires indicators to be directly comparable. In the following analysis, we apply six criteria to assess these qualitative characteristics. The validity of different measures is based on their correctness and completeness in covering overall environmental impacts whereby the latter includes a content- and a time-related dimension. Major criteria to assess reliability are quantifiability, objective verifiability and comparability.

Regarding their relationship, validity dominates reliability and thus needs to be examined first.[10] Which measurement categories provide a valid depiction of EP? When comparing the dimensions of EP according to the model of Wood (1991) to our classification of EP indicators, outcomes and impacts of environmental responsibility refer to operational EP measures while principles of environmental responsibility

[9] In their assessment of how well a company fulfilled their expectations, stakeholders might also preferably rely on reliable indicators. Furthermore, their access to information is limited. Thus, information availability and reliability moderate the relationship of EP and stakeholders' assessments.

[10] Data quality is irrelevant if it is the *wrong* measure (Kromrey 2009).

EP dimensions	EP indicators	
Structural principles of environmental responsibility	Strategic	Indicators of environmental attitudes and objectives
Processes of environmental responsiveness		Indicators of strategic environmental management
Outcomes of environmental responsibility	Operational	Process oriented indicators
		Input oriented indicators
		Output oriented indicators
Impacts of environmental responsibility		Outcome oriented indicators

Fig. 3 EP dimensions and indicators

and processes of environmental responsiveness are mainly related to strategic EP indicators (see Fig. 3).[11]

Impacts have been identified to not fully match the basic EP definition as they are often influenced by factors beyond the control of companies (see Sect. 3.2). Thus, it is better to measure the performance directly at the company level. Corresponding measurement categories provide a correct depiction of EP. These are:

- operational input indicators
- operational output indicators
- operational process indicators
- indicators of strategic environmental management
- indicators of environmental attitudes and objectives

To correctly measure the success of corporate actions aimed at reducing environmental impacts, indicators should "focus as nearly on outcomes as possible" (Carroll 2000, p. 473). The sources of environmental impacts on the company level are direct interactions with the natural environment through energy and resource consumption (renewable and non-renewable), water consumption, land use, waste generation, and air, soil, and water-polluting emissions (James 1994; Jung et al. 2001; Wood 2010). These indicators correspond to the category of operational inputs and outputs which hence is considered most valid to EP measurement.

In order to ensure a complete depiction of the EP construct, an ideal measure includes *all* components of a company's environmental interactions, weights them in relation to their harmful impacts and combines them to a single indicator (Spicer 1978;

[11] What is called "outcomes" in our categorization as it is the outcome of changes in inputs, operational processes and outputs is labeled as "impacts" by Wood (1991) as it impacts the company and its stakeholders. By contrast, Wood's "outcomes" refer to the outcomes of principles and processes. Wood (2010) also lists measures that cannot be assigned to a single dimension. What she calls "multi-category measures" are "combined measures" in our categorization.

Ruf et al. 1998; Tyteca 2002). In this respect, using single input and output indicators to proxy EP is criticized to not reflect overall EP properly (see Sect. 2.3). However, measuring EP by *multiple* input and output indicators leads to further problems. Each company has numerous interactions with the natural environment which are difficult to cover in total. In addition, each company has a unique profile of physical inflows and outflows, depending on its specific products and business processes. Comparability of input and output levels is only given within peers, e.g., sub-industries using the same technologies (Telle 2006). E.g., the water consumption is different for the production of milk and beer, even though both belong to the beverage industry. What might be a good value for one of them would not necessarily apply to the other. If the sample includes milk farmers both from conventional and organic farming, a further separation is required. Within a peer group, performance can be calculated by comparing input and output levels to a benchmark (e.g., target value, mean value). Comparing indicators between different peer groups is only possible on the basis of relative performance levels, e.g. scores or ratings. Large-scale studies are faced with the problem to adequately adjust input and output indicators to each peer group included in their sample.

A third problem is how to weight single indicators. Their weight is determined by their importance in causing environmental impacts. However, capturing environmental impacts leads to a difficult valuation problem whose parameters (e.g., local conditions, detailed environmental impacts of emissions, etc.) are not fully understood so far (Lankoski 2000; Mitnick 2000; Tyteca 2002). Today, the weight attached to single indicators is highly subjective. Overall, input and output indicators suffer from various shortcomings. However, their deficiencies are widely accepted as these indicators convince through data availability, quantifiability and, as they can be observed directly (in contrast to strategic indicators such as attitudes), also through a certain degree of external verifiability.

What has not yet been considered in literature is that completely depicting environmental impacts also has a time-reference. We found stakeholders to be not only interested in a company's current but also in its *future* environmental impacts. Input and output indicators are retrospective and provide a limited view on companies' future environmental impacts what is a major weakness in this respect. Here, process indicators as the second category of operational EP measures might be more important to stakeholders as they set standards for the generation of future outcomes. E.g., in a nuclear reactor, technical precautions and high security standards indicate future performance better than its past inputs and outputs, especially if there never have been any irregular outputs. Nevertheless, process indicators only refer to a short time horizon. How can long-term performance be operationalized? According to Wood's model, operational outcomes are predetermined by a company's principles of environmental responsibility and processes of environmental responsiveness. Corresponding measures, even if they are only indirectly related to environmental impacts through their effects on operational outcomes, indicate the level of attention a company pays to environmental responsibility and the integration of environmental issues into management processes (Klassen and McLaughlin 1996; Banerjee 2002). A company feeling responsible for limiting its environmental damage is likely to initiate adequate actions in the future and to lower its environmental

impacts on a long-term basis. Thus, these dimensions of EP represent the expectations of stakeholders demanding long-term performance.

Regarding the reliability of strategic EP indicators, standardized metrics are hard to find. As information disclosed by the companies lacks comparability, data collection primarily relies on surveys (Tyteca 2002). Measurability is limited as strategic indicators mostly represent qualitative information (Wood 1991; Jung et al. 2001) and a social desirability bias cannot be completely excluded (Banerjee 2002). At least concerning their comparability, strategic indicators are superior to operational indicators. As they refer to general attitudes, management structures and strategic processes that exist in almost every firm, direct comparability of their level is often given even between firms of different industries.

Considering the low reliability of strategic indicators, in particular, their inability to exactly quantify the level of environmental impacts and their lack of external verifiability, it could be argued that forecasts of future environmental impacts based on retrospective operational EP indicators are more precise. But there is a further shortcoming that should be taken into account. As disclosure is mostly voluntary, many companies only provide beneficial rather than accurate, unbiased information on their current performance. Stakeholders are confronted with the problem of figuring out a company's "true" efforts to reduce environmental impacts. One way is to look for signs of negative outcomes beyond the direct control of the companies, e.g., information required by law or negative environmental events such as accidents or crises. This could explain the findings of event studies that indicators of a "bad" EP consistently cause negative financial effects (see Sect. 2.1 and Wood 2010). This is also closely connected with the observation that stakeholders view environmental activities less favorably when a link to profitability is apparent (McWilliams et al. 2006). We argue that the main problem is not that stakeholders condemn companies of earning money when being green. If they perceive a company's behavior to be credible and trustworthy, they regard a high CP with favor. The reason for negative reactions rather lies in the fact that stakeholders search for credible indicators separating good performers from the black sheep. A coherence of EP and profit targets might be regarded as an indicator of greenwashing behavior. If stakeholders perceive responsible behavior to be conducted to earn money, they might come to the conclusion that true EP is low and even react negatively to companies' environmental activities. Transferred to the context of EP measurement in empirical research, operational process measures and strategic EP measures might score better than input and output indicators as they usually provide a broad coverage of the EP construct (e.g., EMS certifications, environmental attitudes) and thus guarantee a certain degree of overall performance. If they can be classified as reliable, they can be used to assess the risk of unrecognized greenwashing behavior.

In summarizing, our analysis of validity and reliability (operationalized by the attributes of a correct depiction and a content- as well time-based completeness in covering the construct validity as well as quantifiability, objective verifiability and comparability for reliability) of different categories of EP measures leads to mixed results (see Fig. 4).

As can be seen, all categories vary substantially in their degree of validity and reliability. As operational input and output-based measures are most directly connected

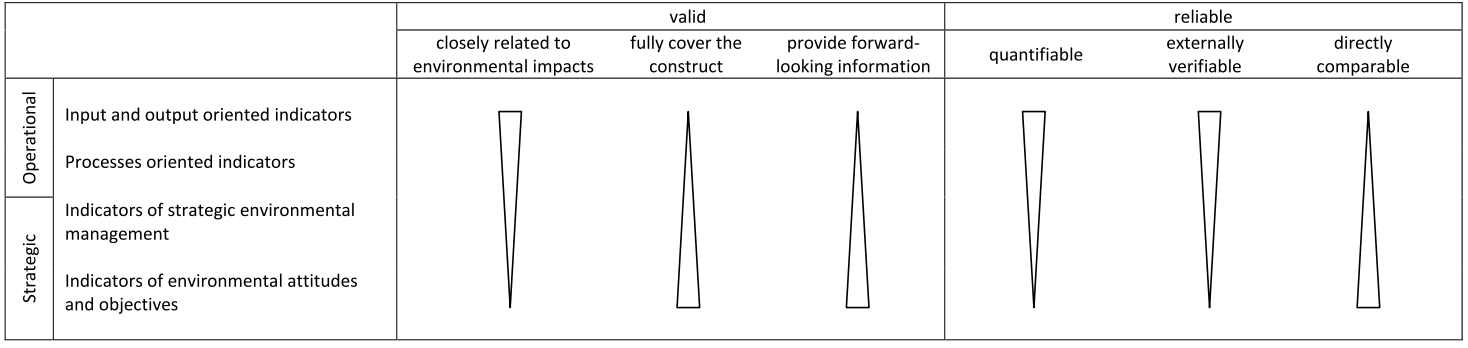

Fig. 4 Qualitative characteristics of EP indicators

to a company's environmental impacts, they allow (more or less) precise conclusions on the level of single environmental impacts. However, input and output measures refer to a multitude of environmental interactions that can hardly be covered in total. Furthermore, they only deliver retrospective information. Operational process indicators and strategic EP measures perform considerably better in this respect. They provide forward-looking indicators and yield a broader view of a company's EP that helps to assess overall EP. Usually they also allow direct comparisons between different firms, while specific input and output indicators can only be compared within small peer groups. However, strategic indicators quantify the level of environmental impacts less exactly and can hardly be verified externally.

We conclude from these results that indicators of operational inputs and outputs, operational processes as well as strategic EP are all highly relevant to EP measurement as stakeholders' expectations potentially refer to all of them. A combination of several categories would reflect stakeholders' expectations better than focusing on just one category, e.g., by measuring a handful of input and output indicators. One approach would be to measure EP by a scoring model that weights the different issues according to their relevance to stakeholders. Researchers should also pay more attention to the reliability of the measures they use, e.g., to compare specific indicators only within the relevant peer-group.

5 Conclusion

Current empirical studies examining the economic impacts of EP mainly focus on moderating and influencing factors of the relationship by asking "When and why does it pay to be green?". In these studies, EP is measured by various operational, strategic or combined measures. Their heterogeneity and missing theoretical foundation are seen as major reasons of the inconsistent empirical results within this topic. It has been recommended to solve measurement problems before further conducting empirical research. In particular, this should be done by reconsidering theoretical fundaments and refocusing on stakeholders.

By applying the framework of Wood (1991) to the EP construct, we identify three dimensions to which EP measures can refer. However, we exclude indicators not directly measuring EP at the company level. By this, we identify five measurement categories directly corresponding to the EP construct: operational input indicators, output indicators, process indicators, indicators of strategic environmental management and indicators of environmental attitudes and objectives. Measures belonging to these categories (or combinations thereof) potentially provide construct validity.

To further operationalize EP, we rely on the basic definition of EP (also underlying the model of Wood) as the fulfillment of stakeholders' environmental expectations and examine those expectations in detail. Our analysis reveals that there is not one best way to measure EP but—depending on the specific research question and the sample—the operationalization of EP should adapt to (1) special interests of the stakeholder groups being examined and their time horizon, (2) special characteristics of companies/products and (3) external factors relevant to the expectations of stakeholders.

By analyzing the validity and reliability of different categories of EP measures within large-scale studies, we show that operational input and output indicators, even though they are closely connected to the companies' environmental impacts, have a number of shortcomings that in prior literature have not been taken sufficiently into account. In particular, input and output oriented indicators only provide a retrospective view. Operational process indicators and strategic EP indicators are able to provide a better assessment of future environmental impacts. They also provide a broader coverage of the EP construct, i.e., they rather measure overall performance than, as input and output indicators, to depict only small aspects of EP. Regarding reliability, all measures have strengths and weaknesses. In particular, input and output indicators are only comparable within narrow groups. In contrast, strategic indicators such as environmental responsible attitudes can be compared more easily across different companies, what makes them comfortable to use in large-scale studies.

What does that mean for the measurement of EP? Researchers, in order to obtain consistent results, should closely examine the environmental expectations of stakeholders within the context of their study and adjust the choice of indicators as precise as possible. In contrast, developing generic EP measures means to find a compromise between different stakeholder expectations and thus is related to loss in validity. At least, they should represent a *balanced* picture of stakeholders' expectations. We showed that this does not only include operational input and output indicators but also other categories, e.g., operational process measures and strategic measures. In this context, practical considerations today dominating the choice of EP measures, e.g., data availability, should be on the second position.

Our results open several avenues for further research. One issue is to improve knowledge on stakeholders' environmental expectations to better understand the extent of EP. Up to date, they have barely been analyzed empirically.

A crucial step in developing standardized EP measures on the basis of companies' environmental impacts is to enhance the theoretical understanding of indicators that grasp different components of the complex construct. While our analysis focused on basic characteristics of coherent measures that can be used to assess EP, further research should deepen this analysis and provide guidance on the selection of specific indicators. Also the weighting and aggregation of different indicators according to their relative importance within comparable groups requires further examination.

A third point is related to the distinction of EP from related constructs as stakeholders' perceptions, evaluations and actions which moderate the relationship of EP and CP. Based on these insights, a reclassification of empirical studies according to the constructs that were actually measured would provide valuable insights into the relationship of EP and CP.

Our paper has several limitations. Notably, not only stakeholder induced ("indirect") effects determine the economic benefits of environmental protection activities. However, our analysis did not account for direct effects of environmental protection activities, e.g., savings from a more efficient use of resources and a lower appearance of waste and toxic emissions (Günther et al. 2004; Weber 2008). Direct effects are relatively easy to calculate and thus little noticed in empirical research. But as they also contribute to the economic impacts of EP, broadly applicable EP measures should cover them, too. However, we did not answer the question if our results are applicable to capture direct effects.

Based on our findings, we contribute to the development of coherent EP measures. Even though EP measurement today is far from being consistent, our results provide some basic guidance on this topic. By reducing the multitude of EP measures to a smaller number of well-founded measures the consistency of empirical results and for this reason also the research on economic consequences of environmental protection could be considerably enhanced.

Appendix: Summary of empirical studies

Study	Time period and sample	EP indicators	CP indicators	Other variables (Type: MED = mediator variable; MOD = moderator variable; CON = control variable)	Results
Regression analyses					
Hart and Ahuja (1996)	1989–1992; 127 publicly traded manufacturing, mining and production firms	Emissions efficiency index	ROS, ROA, ROE	CON: R&D intensity, capital intensity, advertising intensity, leverage, industry average performance	Emissions reduction enhanced CP in time period $t + 1$ and even more in $t + 2$, impact was stronger for firms with high pollution level
Nehrt (1996)	Mid 1980's-early 1990's; 50 chemical bleached paper pulp manufacturers of 8 countries	Timing of investment (years of investment in pollution control technology), intensity of investment (percentage of chlorine-free pulp production)	Net income growth	CON: timing of pollution control regulations, GDP growth, wages growth, initial net income, sales growth	Timing of investment is positively and intensity of investment negatively correlated with net income growth
Cordeiro and Sarkis (1997)	1992; 523 US firms	Environmental proactivism (emissions and waste releases), changes in proactivism	Industry analyst 1- and 5-year EPS performance forecasts	CON: firm size, leverage	Significant negative relationship between environmental proactivism and EPS forecasts
Griffin and Mahon (1997)	1990 and 1992; 7 large US chemical firms	KLD rating, FAMA index, emission data, corporate philanthropy activities	ROS, ROA, ROE, total assets, asset age		FMA and KLD indices track one another, TRI and corporate philanthropy measures do not correlate to CP
Russo and Fouts (1997)	1991–1992; 243 firms	Franklin Research and Development Corporation (FRDC) rating	ROA	MOD: industry growth CON: industry concentration, firm growth, firm size, capital intensity, advertising intensity	FRDC rating is positively correlated with ROA, industry growth strengthens the relationship

Study	Time period and sample	EP indicators	CP indicators	Other variables (Type: MED = mediator variable; MOD = moderator variable; CON = control variable)	Results
Waddock and Graves (1997)	1990–1991; 469 publicly traded firms	KLD rating	ROS, ROA, ROE	CON: company size, risk, industry	Better KLD rating leads to improved CP in the following year
Judge and Douglas (1998)	1992–1994; 196 US firms	Perceived EP (e.g., compliance with environmental regulations, proactive environmental management activities)	Perceived CP, ROA, ROS	MOD: integration of environmental issues into the strategic planning process CON: firm size	Integration of environmental issues capability improved both EP and CP indicators
Bhat (1999)	1988–1990; 230 US firms	Waste discharged into air, land and water in pounds per sales revenue	Profit margins, stock market values	CON: R&D expenditures, advertisement expenditures	Lower pollution is correlated with higher profit margin and stock market values
Christmann (2000)	1992–1994; 88 chemical firms	"Best practices" of environmental management (pollution prevention, innovation of pollution prevention technologies, early timing of environmental strategies)	Perceived cost advantage	MOD: complementary assets CON: business unit size, environmental issue characteristics	No significant cost effects of pollution-prevention and early timing; innovation of proprietary technologies is positively correlated with cost advantages; complementary assets have a positive influence
McWilliams and Siegel (2000)	1991–1996; 524 publicly traded firms	KLD rating	ROA	CON: R&D intensity, industry	When control variables are included, the previously positive correlation of KLD ratings and ROA gets insignificant
Dowell et al. (2000)	1994–1997; 89 manufacturing and mining firms	Adoption of global environmental standards	Tobin's q	CON: R&D intensity, advertising intensity, leverage, multinationality, firm size, industry	Firms adopting global standards have higher market values

Study	Time period and sample	EP indicators	CP indicators	Other variables (Type: MED = mediator variable; MOD = moderator variable; CON = control variable)	Results
Karagozoglu and Lindell (2000)	89 high tech and manufacturing firms	Proactiveness of environmental strategy (ES) (recycling, resource use, energy efficiency, use of hazardous substances, pollution-prevention)	Perceived financial performance (profit margin, market share, overall CP)	MED: environmental innovativeness (EI), environmental competitive advantage (ECA) MOD: regulatory stringency (SRS), regulatory supportiveness (RS) CON: firm size, industry	The correlation of ES and CP was not significant but is mediated by EI and ECA; effect of RS and SRS on the correlation of ES and EI was not significant; RS strengthened the positive correlation of EI and ECA
Álvarez Gil et al. (2001)	1998; 294 Spanish hotels	Environmental management (e.g., environmental costs and savings, training programs, purchasing policies, energy- and water-saving actions, recycling)	Perceived CP (current-year profitability, profitability over the last 3 years, occupancy rate)		Environmentally proactivity is accompanied by a higher CP level
King and Lenox (2001)	1987–1996; 652 US manufacturing firms	Total emissions, industry emissions, relative emissions of facilities	Tobin's q	CON: firm size, capital intensity, firm growth, leverage, R&D intensity, regulatory stringency, permits	Emissions are significantly negatively related to CP
Konar and Cohen (2001)	1989 (partly lagged); 321 firms	Pounds of toxic chemical emissions to revenue, number of environmental lawsuits	Tobin's q, intangible asset value	CON: market share, industry concentration ratio, sales growth, advertising intensity, R&D intensity, firm size, import intensity	Both toxic emissions and environmental lawsuits are negatively correlated with CP

Study	Time period and sample	EP indicators	CP indicators	Other variables (Type: MED = mediator variable; MOD = moderator variable; CON = control variable)	Results
King and Lenox (2002)	1991–1996; 614 manufacturing firms	Total emissions	ROA, Tobin's q	MED: waste prevention, waste treatment, waste transfer CON: firm size, firm growth, capital intensity, leverage, R&D intensity, regional wages, regulatory stringency, permits	Reduced total emissions enhance CP, this is mediated by the waste control method
Wagner et al. (2002)	1995–1997, 37 pulp and paper manufacturers of 4 countries	Environmental index based on SO_2, NO_x and COD emissions per production unit	ROS, ROE, ROCE	CON: country, firm size, capital intensity, financial leverage, sub-sector-membership	Emissions have a positive impact on ROCE but not on ROS and ROE
Ziegler et al. (2002)	1996–2001; 214 European firms	Environmental and social sustainability of industry sector, environmental and social sustainability of firm in relation to competitors (based on Sarasin & Cie ranking)	Average monthly stock returns	CON: market capitalization, market-to-book ratio, country, industry	Environmental sustainability of industry sector is positively and social sustainability of industry sector negatively correlated with stock returns; other measures show no significant results
Al-Tuwaijri et al. (2004)	1994; 198 firms	Ratio of toxic waste recycled to total toxic waste generated	Annual industry-adjusted stock return	CON: unexpected earnings, predisclosure environment, growth opportunities, profit margin, environmental exposure, environmental concern, public visibility, firm size	Significantly positive relation between waste recycling and stock returns

Study	Time period and sample	EP indicators	CP indicators	Other variables (Type: MED = mediator variable; MOD = moderator variable; CON = control variable)	Results
Melnyk et al. (2003)	1222 business units of 5 industries	State of the environmental management system	Perceived operational performance indicators (e.g., overall costs, product quality), environmental options (e.g., product and process redesign, disassembly , alliances)	CON: age of the EMS; firm size, ownership	The presence of a formal or certified EMS is positively correlated with CP indicators and environmental options
Wagner (2005b)	1995–1997, 37 pulp and paper manufacturers of 4 countries	Emission-based index based on SO_2, NO_x and COD emissions per production unit, input-based index based on total energy input and total water input per production unit	ROS, ROE, ROCE	MOD: environmental strategy type CON: country, capital intensity, financial leverage, sub-sector-membership, firm size	The emission-based index is negatively related to CP; the input-based index is not significant; pollution prevention-oriented strategies influence the relationship between EP and CP positively
Brammer and Pavelin (2006)	1998–2002, 210 publicly traded UK firms	Score based on community, environmental (policies, systems, reporting, impacts) and employee performance (e.g., health and safety, employee relations, job security, job creation)	BMAC index	MOD: industry sectors, different types of CSP CON: financial performance, market risk, leverage, firm size, media exposure, R&D intensity, advertising intensity, ownership	Positive effect of CSP on CP is restricted to some sectors and to community performance; environmental and employee performance become significant when varying across sectors is permitted, environmental performance harms reputation in some sectors and in others it supports reputation

Study	Time period and sample	EP indicators	CP indicators	Other variables (Type: MED = mediator variable; MOD = moderator variable; CON = control variable)	Results
Luo and Bhattacharya (2006)	2001–2004; 113 firms	FAMA index	Tobin's q, stock return	MED: customer satisfaction (CS) MOD: innovativeness capability (IN), product quality (PQ) CON: ROA, firm size, R&D intensity, competition intensity, advertising intensity	The FAMA index affects CP positively; the relationship of FAMA and CP is partly mediated by CS; if IN is low, FAMA reduces CS and harms market value
Sammer and Wüstenhagen (2006)	2004; 151 Swiss consumers	Eco-labeling of washing machines (water and energy consumption, energy efficiency)	Consumer choice	CON: brand, price, equipment	Water and energy consumption and energy efficiency are positively correlated with consumer choice
Telle (2006)	1990–2001; 85 Norwegian plants of 4 industries	Index consisting of greenhouse gas, acid, NMVOC-equivalents and particle emissions per production unit	ROS	MOD: industry CON: regulatory burden, firm size, year, plant heterogeneity	When controlling for plant heterogeneity, the previously positive correlation of low emission and ROS gets insignificant; industry moderates the relation of emission and ROS
Dick et al. (2008)	2000–2005; 7.500 Spanish firms	ISO 14001 certification	ROA, sales growth		Certification is associated with a higher ROA and sales growth, but CP is already increased before certification
Busch and Hoffmann (2011)	2006–2007; 174 firms of 9 industries	Carbon intensity (ratio between total GHG emissions and sales), carbon management (score based on carbon management targets and strategies)	ROA, ROE, Tobin's q	MOD: carbon management CON: company size, financial risk, country, industry	A reduced carbon intensity is positively related with Tobin's q; carbon management has a negative impact on all CP indicators and does not moderate the correlation of carbon intensity and CP

Study	Time period and sample	EP indicators	CP indicators	Other variables (Type: MED = mediator variable; MOD = moderator variable; CON = control variable)	Results
Clarkson et al. (2011)	1990–2003, 242 US firms of 4 industry sectors	Inverse of pollution propensity (toxics releases in pounds scaled by the cost of goods sold)	ROA, CF, enterprise value	MOD: management talent or capability CON: firm size, age of equipment, capital intensity	Pollution reduction leads to economic benefits in subsequent years
Event studies					
Blacconiere and Patten (1994)	1984; 47 chemical firms	Union Carbide's chemical leak in Bhopal, India in 1984	Daily abnormal returns	MOD: significance of involvement in the chemical industry; the extent of environmental disclosure CON: firm size	A significant negative industry market reaction occurred; this was enhanced by a high involvement in the chemical industry and weakened by environmental disclosure
Hamilton (1995)	1989; 436 firms	Release of TRI emission data	Daily abnormal returns, media coverage	MOD: emission type, Superfund sites CON: ownership, firm size, industry	High polluters also to face negative abnormal returns, they were higher for firms with Superfund sites; media coverage increased for high-polluters, particularly regarding air releases and offsite shipments
Klassen and McLaughlin (1996)	1985–1991; 96/16 firms	Announcements of environmental awards, environmental crisis	Daily abnormal returns, emissions, compliance, overall environmental rating, systematic risk	MOD: industry, first time effects, award-type, environmental expenditures of industry CON: market, firm size	Awards announcements are followed by positive and crises by negative abnormal returns; awards are associated with lower emissions and a better environmental rating; announcements decrease equity beta; industry and first time effects influence abnormal returns
Halme and Niskanen (2001)	1970–1996, 10 Finnish forest industry firms	News on environmental investments	Daily abnormal returns		Stock prices declined significantly after the news emerged but recovered fully after one week

Study	Time period and sample	EP indicators	CP indicators	Other variables (Type: MED = mediator variable; MOD = moderator variable; CON = control variable)	Results
Portfolio analyses					
Hamilton et al. (1993)	1981–1990; 32 funds	Socially responsible mutual funds	Excess returns compared to NYSE and conventional mutual funds	MOD: year of establishment	Socially responsible mutual funds do not earn excess returns, establishment year has no significant effect
White (1996)	1989–1992; 97 publicly traded firms	CEP rating	Jensen's alpha	MOD: adoption of CERES principles	A portfolio of firms with high environmental reputation has higher risk-adjusted returns than the overall market or other portfolios; decision to adopt CERES principles leads to positive abnormal returns
Statman (2000)	1990–1998; 32 funds	Domini Social Equity Fund (based on DSI), other socially responsible mutual funds	Nominal and risk-adjusted returns, Jensen's alpha, excess standard-deviation-adjusted return		The DSI did as well as the S& P500, other socially responsible mutual funds did worse than DSI and S& P500 but no worse than conventional mutual funds
Bauer et al. (2005)	1990–2001; 103 German, UK and US funds	Ethical mutual funds	Jensen's alpha, 4-factor alpha	MED: market risk, caps size MOD: country, time	No significant performance difference between ethical and conventional mutual funds; difference varies over time; ethical funds are less exposed to market risk and have a higher growth stock orientation; German and UK funds are more invested in small and US funds more in big caps; ethical funds are not able to out-perform ethical equity indices

References

Adamczyk, S., Hansen, E. G., & Reichwald, R. (2009) Measuring sustainability by environmental innovativeness: results from action research at a multinational corporation in Germany. Paper presented at the International Conference on Evaluation Metrics of Corporate Social and Environmental Responsibility 8–10 Jun, Lyon.

Agle, B., & Kelley, P. (2001). Ensuring validity in the measurement of corporate social performance: lessons from corporate United Way and PAC campaigns. *Journal of Business Ethics*, *31*, 271–284.

Allouche, J., & Laroche, P. (2005). A meta-analytical investigation of the relationship between corporate social and financial performance. *Revue de Gestion Des Ressources Humaines*, *57*, 18–41.

Al-Tuwaijri, S. A., Christensen, T. E., & Hughes, K. E. II (2004). The relations among environmental disclosure, environmental performance, and economic performance: A simultaneous equations approach. *Accounting, Organizations and Society*, *29*, 447–471.

Álvarez Gil, M., Burgos Jiménez, J., & Céspedes Lorente, J. (2001). An analysis of environmental management, organizational context and performance of Spanish hotels. *Omega*, *29*, 457–471.

Ambec, S., & Lanoie, P. (2007) When and why does it pay to be green? Cahier de recherche no. IEA-07-04, HEC Montréal.

Ambec, S., & Lanoie, P. (2008). Does it pay to be green? A systematic overview. *Academy of Management Perspectives*, *22*, 45–62.

Atkinson, A. A., Waterhouse, J. H., & Wells, R. B. (1997). A stakeholder approach to strategic performance measurement. *MIT Sloan Management Review*, *38*, 25–37.

Azzone, G., & Noci, G. (1996). Measuring the environmental performance of new products an integrated approach. *International Journal of Production Research*, *34*, 3055–3078.

Banerjee, S. B. (2002). Corporate environmentalism. The construct and its measurement. *Journal of Business Research*, *55*, 177–191.

Banerjee, S. B., Iyer, E. S., & Kashyap, R. K. (2003). Corporate environmentalism: antecedents and influence of industry type. *Journal of Marketing*, *67*, 106–122.

Bartolomeo, M. (1995). Environmental performance indicators in industry. Nota di lavoro 41/95. Fondazione ENI Enrico Mattei, Milan.

Bauer, R., Koedijk, K., & Otten, R. (2005). International evidence on ethical mutual fund performance and investment style. *Journal of Banking & Finance*, *29*, 1751–1767.

Berkhout, F., Azzone, G., Carlens, J., et al. (2001). Measuring the environmental performance of industry (MEPI). Final report. University of Sussex, Brighton.

Bhat, V. N. (1999). Does it pay to be green? *International Journal of Environmental Studies*, *56*, 497–507.

Biondi, V., Frey, M., & Iraldo, F. (2000). Environmental management systems and SMEs. Motivations, opportunities and barriers related to EMAs and ISO 14001 implementation. *Greener Management International*, *29*, 55–69.

Blacconiere, W. G., & Patten, D. M. (1994). Environmental disclosures, regulatory costs, and changes in firm value. *Journal of Accounting & Economics*, *18*, 357–377.

Bloom, P. N., Hoeffler, S., Keller, K. L., & Basurto Meza, C. E. (2006). How social case marketing affects consumer perceptions. *MIT Sloan Management Review*, *47*, 49–55.

Brammer, S. J., & Pavelin, S. (2006). Corporate reputation and social performance. The importance of fit. *Journal of Management Studies*, *43*, 435–455.

Brewer, G. D., & Stern, P. C. (2005). *Decision making for the environment. Social and behavioral science research priorities*. Washington: National Academies Press.

Busch, T., & Hoffmann, V. H. (2011). How hot is your bottom line? Linking carbon and financial performance. *Business & Society*, *50*, 233–265.

Carroll, A. B. (2000). A commentary and an overview of key questions on corporate social performance measurement. *Business & Society*, *39*, 466–478.

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

Christmann, P., & Taylor, G. (2001). Globalization and the environment. Strategies for international voluntary environmental initiatives. *Journal of International Business Studies*, *32*, 439–458.

Clarkson, P. M., Li, Y., Richardson, G. D., & Vasvari, F. P. (2008). Revisiting the relation between environmental performance and environmental disclosure. An empirical analysis. *Accounting, Organizations and Society*, *33*, 303–327.

Clarkson, P. M., Li, Y., Richardson, G. D., & Vasvari, F. P. (2011). Does it really pay to be green? Determinants and consequences of proactive environmental strategies. *Journal of Accounting and Public Policy*, *30*, 122–144.

Cordeiro, J. J., & Sarkis, J. (1997). Environmental proactivism and firm performance. Evidence from security analyst earnings forecasts. *Business Strategy and the Environment*, *6*, 104–114.

Dick, G., Heras, I., & Molina-Azorín, J. F. (2008). Cause and effect? ISO 14001 certification and financial benefits. Working paper No. 168, Kent Business School.

Dixon-Fowler, H. R., Slater, D. J., Romi, A.M., et al. (2009). Beyond "does it pay to be green?". A meta-analysis of moderators oft the CEP and CFP Relationship. Paper presented at the Academy of Management Annual Meeting 6–11 Aug, Chicago.

Dowell, G., Hart, S., & Yeung, B. (2000). Do global environmental standards create or destroy market value? *Management Science*, *46*, 1059–1074.

Du, S., Bhattacharya, C. B., & Sen, S. (2010). Maximizing business returns to corporate social responsibility (CSR). The role of CSR communication. *International Journal of Management Reviews* , *12*, 8–19.

Freeman, R. E. (1984). *Strategic management. A stakeholder approach*. Boston: Pitman.

Griffin, J. J., & Mahon, J. F. (1997). The corporate social performance and corporate financial performance debate. *Business & Society*, *36*, 5–31.

Günther, E., Günther, T., & Hoppe, H. (2004). Are environmental aspects value drivers for companies? A review of empirical studies. Dresdner Beiträge zur Betriebswirtschaftslehre Nr. 81/04. TU Dresden.

Halme, M., & Niskanen, J. (2001). Does corporate environmental protection increase or decrease shareholder value? The case of environmental investments. *Business Strategy and the Environment*, *10*, 200–214.

Hamilton, J. T. (1995). Pollution as news. Media and stock market reactions to the toxics release inventory data. *Journal of Environmental Economics and Management*, *28*, 98–113.

Hamilton, S., Jo, H., & Statman, M. (1993). Doing well while doing good? The investment performance of socially responsible mutual funds. *Financial Analysts Journal*, *49*, 62–66.

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

Ilinitch, A. Y., Soderstorm, N. S., & Thomas, T. E. (1998). Measuring corporate environmental performance. *Journal of Accounting and Public Policy*, *17*, 383–408.

James, P. (1994). Business environmental performance measurement. *Business Strategy and the Environment*, *3*, 59–67.

Judge, W. Q., & Douglas, T. J. (1998). Performance implications of incorporating natural environmental issues into the strategic planning process. An empirical assessment. *Journal of Management Studies*, *35*, 241–262.

Jung, E. J., Kim, J. S., & Rhee, S. K. (2001). The measurement of corporate environmental performance and its application to the analysis of efficiency in oil industry. *Journal of Cleaner Production*, *9*, 551–563.

Karagozoglu, N., & Lindell, M. (2000). Environmental management. Testing the win-win model. *Journal of Environmental Planning and Management*, *43*, 817–829.

King, A., & Lenox, M. (2001). Does it really pay to be green? An empirical study of firm environmental and financial performance. *Journal of Industrial Ecology*, *5*, 105–116.

King, A., & Lenox, M. (2002). Exploring the locus of profitable pollution reduction. *Management Science*, *48*, 289–299.

Klassen, R. D., & McLaughlin, C. P. (1996). The impact of environmental management on firm performance. *Management Science*, *42*, 1199–1214.

Konar, S., & Cohen, M. A. (2001). Does the market value environmental performance? *Review of Economics and Statistics*, *83*, 281–289.

Kromrey, H. (2009). *Empirische Sozialforschung* (12th edn.). Stuttgart: UTB.

Lankoski, L. (2008). Corporate responsibility activities and economic performance. A theory of why and how they are connected. *Business Strategy and the Environment*, *17*, 536–547.

Lankoski, L. (2000). Determinants of environmental profit. An analysis of the firmlevel relationship between environmental performance and economic performance. Dissertation, Helsinki. University of Technology.

Luo, X., & Bhattacharya, C. B. (2006). Corporate social responsibility, customer satisfaction, and market value. *Journal of Marketing*, *70*, 1–18.

Marom, I. Y. (2006). Toward a unified theory of the CSP–CFP link. *Journal of Business Ethics*, *67*, 191–200.

McGuire, J. B., Sundgren, A., & Schneeweis, T. (1988). Corporate social responsibility and firm financial performance. *Academy of Management Journal*, *31*, 854–872.

McWilliams, A., & Siegel, D. (2000). Corporate social responsibility and financial performance. Correlation or misspecification? *Strategic Management Journal*, *21*, 603–609.
McWilliams, A., Siegel, D. S., & Wright, P. M. (2006). Corporate social responsibility. Strategic implications. *Journal of Management Studies*, *43*, 1–18.
Melnyk, S. A., Sroufe, R. P., & Calantone, R. (2003). Assessing the impact of environmental management systems on corporate and environmental performance. *Journal of Operations Management*, *21*, 329–351.
Miles, M. P., & Covin, J. G. (2000). Environmental marketing. A source of reputational, competitive and financial advantage. *Journal of Business Ethics*, *23*, 299–311.
Milgrom, P. R., & Roberts, J. (1992). *Economics, organization and management*. Englewood Cliffs: Prentice Hall.
Mitnick, B. M. (2000). Commitment, revelation, and the testaments of belief: the metrics of measurement of corporate social performance. *Business & Society*, *39*, 419–465.
Molina-Azorín, J. F., Claver-Cortés, E., López-Gamero, M. D., & Tarí, J. J. (2009). Green management and financial performance. A literature review. *Management Decision*, *47*, 1080–1100.
Nehrt, C. (1996). Timing and intensity effects in environmental investments. *Strategic Management Journal*, *17*, 535–547.
Orlitzky, M., Schmidt, F. L., & Rynes, S. L. (2003). Corporate social and financial performance: a meta-analysis. *Organization Studies*, *24*, 403–441.
Orlitzky, M., Siegel, D. S., & Waldman, D. A. (2011). Strategic corporate social responsibility and environmental sustainability. *Business & Society*, *50*, 6–27.
Pava, M. L., & Krausz, J. (1996). The association between corporate social-responsibility and financial performance. The paradox of social cost. *Journal of Business Ethics*, *15*, 321–357.
Peloza, J., & Papania, L. (2008). The missing link between corporate social responsibility and financial performance. Stakeholder salience and identification. *Corporate Reputation Review*, *11*, 169–181.
Promberger, K., & Spiess, H. (2006). Der Einfluss von corporate social (and ecological) responsibility auf den Unternehmenserfolg. Working paper 26/2006, University of Innsbruck.
Reinhardt, F. L. (1998). Environmental product differentiation. Implications for corporate strategy. *California Management Review*, *40*, 43–73.
Roome, N., & Wijen, F. (2006). Stakeholder power and organizational learning in corporate environmental management. *Organization Studies*, *27*, 235–263.
Ruf, B. M., Muralidhar, K., & Paul, K. (1998). The development of a systematic, aggregate measure of corporate social performance. *Journal of Management*, *24*, 119–133.
Russo, M. V., & Fouts, P. A. (1997). A resource-based perspective on corporate environmental performance and profitability. *Academy of Management Journal*, *40*, 534–559.
Salzmann, O., Ionescu-Somers, A., & Steger, U. (2005). The business case for corporate sustainability. Literature review and research options. *European Management Journal*, *23*, 27–36.
Sammer, K., & Wüstenhagen, R. (2006). The influence of eco-labeling on consumer behaviour. Results of a discrete choice analysis for washing machines. *Business Strategy and the Environment*, *15*, 185–199.
Schaltegger, S., & Müller, K. (1997). Calculating the true profitability of pollution prevention. *Greener Management International*, *17*, 53–68.
Schaltegger, S., & Synnestvedt, T. (2002). The link between 'green' and economic success. Environmental management as the crucial trigger between environmental and economic performance. *Journal of Environmental Management*, *65*, 339–346.
Spicer, B. H. (1978). Investors, corporate social performance and information disclosure. An Empirical Study. *The Accounting Review*, *53*, 94–111.
Statman, M. (2000). Socially responsible mutual funds. *Financial Analysts Journal*, *56*, 30–39.
Sturm, A. (2000). Performance measurement und environmental performance measurement. Dissertation, TU Dresden.
Telle, K. (2006). "It pays to be green". A premature conclusion? *Environmental & Resource Economics*, *35*, 195–220.
Tyteca, D. (2002). Business organisational response to environmental challenges: performance measurement and reporting. Working paper 50/02, IAG School of Management.
Waddock, S. A., & Graves, S. B. (1997). The corporate social performance-financial performance link. *Strategic Management Journal*, *18*, 303–319.
Wagner, G. R. (1997). *Betriebswirtschaftliche Umweltökonomie*. Stuttgart: Lucius & Lucius.
Wagner, M. (2001). A review of empirical studies concerning the relationship between environmental and economic performance. What does the evidence tell us? CSM, Luneburg.

Wagner, M. (2005a). *Consistency and credibility? Environmental reporting, environmental performance indicators and economic performance*. Marburg: Tectum.

Wagner, M. (2005b). How to reconcile environmental and economic performance to improve corporate sustainability. Corporate environmental strategies in the European paper industry. *Journal of Environmental Management*, *76*, 105–118.

Wagner, M., van Phu, N., Azomahou, T., & Wehrmeyer, W. (2002). The relationship between the environmental and economic performance of firms. An empirical analysis of the European paper industry. *Corporate Social Responsibility and Environmental Management*, *9*, 133–146.

Walter, K. (2005). *Wettbewerbsvorteile durch Umweltmanagement. Ökologische Herausforderungen effektiv nutzen*. Berlin: VDM.

Weber, M. (2008). The business case for corporate social responsibility. A company-level measurement approach for CSR. *European Management Journal*, *26*, 247–261.

White, M. A. (1996). *Corporate environmental performance and shareholder value*. Charlotteville: University of Virginia.

Wood, D. J. (1991). Corporate social performance revisited. *The Academy of Management Review*, *16*, 691–718.

Wood, D. J. (2010). Measuring corporate social performance. A review. *International Journal of Management Reviews* , *12*, 50–84.

Wood, D. J., & Jones, R. E. (1995). Stakeholder mismatching. A theoretical problem in empirical research on corporate social performance. *International Journal of Organizational Analysis*, *3*, 229–267.

Ziegler, A., Rennings, K., & Schröder, M. (2002). Der Einfluss ökologischer und sozialer Nachhaltigkeit auf den Shareholder Value europäischer Aktiengesellschaften. Discussion paper No. 02-32, ZEW. Mannheim.