NHH – Norges Handelshøyskole
and
University of Mannheim (Business School)

Transparency Criteria for the Construction of Corporate Sustainability Ratings

An Analysis Based on the Composite Indicator Framework

Master Thesis

Thesis Advisor: Stein Ivar Steinshamn

Author: Yannic Steffan

Bergen / Mannheim, Fall 2014

Abstract

Corporate sustainability ratings or sustainability ratings (SRs) are applied to integrate sustainability in investment decisions. In order to evaluate the contribution to tackling sustainability challenges and in order to apply them, agencies need to disclose information about the construction of the SR. This thesis follows four steps to present transparency criteria for the construction of SRs. First, I show that the composite indicator framework is applicable to SRs. Second, I use this framework to find out how a one-dimensional measure, which represents a multi-dimensional phenomenon like corporate sustainability, is constructed. This identifies the choices that agencies have when constructing an SR. Third, I use these findings to derive 14 elements that need to be described in order to enable the interpretation of an SR. These elements constitute the transparency criteria. Fourth, the application of the transparency criteria is demonstrated by an empirical review of information that is published during the accreditation by the quality standard Arista 3.0. The last part finds that the accredited agencies do not publish sufficient information to enable the SRs' interpretation and that the Arista 3.0 does not achieve the targeted transparency. I conclude that the transparency criteria are useful to analyze whether disclosed information is sufficient to enable the interpretation of SRs.

Keywords: Sustainability Rating, CR Rating, Sustainable Investing, Composite Indicator

Contents

List of Tables

List of Figures

List of Abbreviations

AHP	Analytical Hierarchy Process
Arise	Association for Responsible Investment Services
CO2e	Carbon dioxide equivalent
DALY	Disability Adjusted Life Years
EC-JRC	European Commission Joint Research Center
ESG	Environmental, Social, and Governance
EU	European Union
Eurosif	European Sustainable and Responsible Investment Forum
FAO	Food and Agriculture Organization of the United Nations
GDP	Gross Domestic Product
GRI	Global Reporting Initiative
IPCC	Intergovernmental Panel on Climate Change
NGO	Non governmental organization
OECD	Organization for Economic Co-operation and Development
OECD handbook	OECD Handbook on Constructing Composite Indicators (2008)
p.	Page
q.	Question
RIR	Responsible investment research
SEEA	System of Environmental-Economic Accounting
SR	Sustainability Rating, Corporate Sustainability Rating
UN	United Nations
UNPRI	United Nations Principles for Responsible Investment
WBCSD	World Business Council for Sustainable Development
WRI	World Resources Institute
WTA	Willingness to accept
WTP	Willingness to pay

Key terms

To assist the reader, key terms are defined below.

Commensurability describes the possibility of a causal relationship between two variables with respect to a shared effect.

Compensability implies that indicators of an aggregation are substitutes. The decrease in one indicator can be substituted by an increase in another indicator. The corresponding adjective is compensatory.

Comparability describes the existence of a causal relationship between two variables that has been quantified with respect to a shared effect.

Composite Indicator is an aggregate of dimensions represented by indicators, and weights "on the basis of an underlying model" (OECD, 2008, p. 13). The term *composite index* is used synonymously in the literature.

Concept is a spelled-out version of a notion. Sustainability ratings measure concepts of corporate sustainability, which may differ from one agency to the next.

Dimensions are constituents of the phenomenon. They correspond to attributes or criteria in the decision-making theory. They may have sub-dimensions. At the lowest level, a variable operationalizes each dimension.

Indicators are variables after normalization. They are always quantitative.

Measurement levels are synonymous to measurement scales. I use levels to avoid confusion of scales and scaling. The theory of measurement levels is summarized in appendix 2.

Methodology refers to the collection of methods, decisions, and procedures during the construction of a composite indicator or sustainability rating.

Notions refer to a generally accepted idea or meaning. The generally accepted idea of sustainability is a notion.

Phenomenon refers to what is being measured by a composite indicator.

Range refers to a continuous set of magnitudes that a variable actually adopts, starting with the smallest, and ending with the largest magnitude.

Scaling refers to a homogeneous transformation of degree 1 to accommodate for differences of the object of analysis, e.g. dividing by sales.

Sustainability Rating (SR) is defined as one-dimensional measure that represents a concept of the multi-dimensional notion of corporate sustainability.

Variables are measures that operationalize dimensions. Variables can be quantitative or qualitative.

1 Introduction

A sustainability rating (SR) measures the sustainability of companies so that financial markets can integrate sustainability in their analyses. As the construction of an SR determines its results, the SR's meaning cannot be interpreted without information about its construction. This thesis sets up transparency criteria regarding the construction of SR's one-dimensional measure that represents corporate sustainability. The composite indicator research framework is chosen to understand the construction of one-dimensional measures and to identify elements that influence the results. As these elements are necessary to interpret an SR, the findings constitute transparency criteria. A review of five SR agencies demonstrates the application of these transparency criteria.

1.1 Relevance

The notion of sustainability has influenced global politics for decades while remaining ill-defined. The Brundtland Report (United Nations (UN), 1987) and the Rio Declaration (UN, 1992) were influential in defining key characteristics of sustainability and have received significant political support. Moldan et al. (2012) subsume the notion of sustainability that is propagated by the two documents as follows:

- Sustainability is anthropocentric, i.e. aspects of sustainability matter to humanity directly or indirectly and are not protected for their own sake.
- Sustainability is multi-dimensional including at least the environmental, social, and economic spheres.
- Sustainability has a long-term perspective, typically beyond the life of one generation.

For its application in financial markets, this notion of sustainability is not concrete enough. Traditionally, investors consider quantitative measures of risk, return, and liquidity. SRs consider quantitative and qualitative measures of environmental, social, and governance (ESG) issues to measure sustainability of companies.

Various investors already consider corporate sustainability in their investment decisions. Estimates of corresponding investments vary. The European association of sustainable investment forums (Eurosif, 2014), approximates the European market of so-called Sustainable and Responsible Investments to be at €10 trillion (10^{12}) in assets under management. Despite recent growth, sustainability considerations are not new to the financial market. The United

Nations Environmental Program Finance Initiative was initiated in 1992 to gather the financial community to take on global environmental challenges. It launched a commitment campaign called UN Principles for Responsible Investment (UNPRI). 1,260 asset owners and investment managers signed the six principles by April of 2014. They state that they "will incorporate ESG issues into investment analysis and decision-making processes" (UNPRI, 2014). The current signatories' assets under management add up to US$45 trillion. Even if only parts of the promised sums are actually invested according to sustainability criteria in the following years, the market will continue to grow. This development is widely hoped for, as the financial market is an efficient lever to incentivize tackling sustainability challenges.

At the center of all these developments is the discrimination between 'normal' investments and sustainable/responsible/green investments. All of these terms are ill defined. Moreover, a nonambiguous operationalization is hampered by the fuzzy nature of 'sustainability' itself (Munda, 2005, p. 978). In the last 15-20 years, agencies have offered SRs as one way to operationalize sustainability in financial markets. SRs evaluate the sustainability of companies by deriving a one-dimensional measure that represents corporate sustainability.

When SRs are integrated into investment decisions, they differentiate sustainable from unsustainable, responsible from irresponsible, and green from not green. Subsequently, SRs influence where money is invested. External assessments of companies were found to increase the likelihood of more responsible behavior of companies (Mackenzie, Rees, & Rodionova, 2013). Among other effects, SRs influence a company's capital costs and reputation. In conclusion, the results of SRs matter. To understand the incentives they apply, it is necessary to interpret SRs and find out what they actually measure. Furthermore, the interpretation is a necessary requirement to apply SRs in financial analyses.

There is no 'correct' way of constructing a measure of corporate sustainability. The concept of corporate sustainability differs among the SRs. A basic distinction can be drawn between the economically motivated approach, where sustainability is conceived to influence risk and return, and the normatively motivated approach, where sustainability is a value on its own, in addition to risk, return, and liquidity (Schäfer, Beer, Zenker, & Fernandes, 2006). To find out what an SR measures and how it defines corporate sustainability, information about its construction is necessary. This thesis focuses on the transparency of SR agencies regarding the construction of SRs. As the construction determines the results, the SR's meaning cannot be interpreted without information about its construction.

The quest for transparency has been postulated and acknowledged widely regarding assessments of sustainability (Hardi & Zdan, 1997; OECD, 2008). A survey among 1.000 "sustainability professionals" found that nearly 90% find disclosure of an SR's methodology is an important determinant for its credibility (Sadowski, Whitaker, & Buckingham, 2010). The same publication concludes that "the sauce remains secret" regarding the ingredients of an SR (p. 5). Besides the industry, the lack of transparency has also sparked criticism by academics (Delmas & Blass, 2010; Delmas, Etzion, & Nairn-Birch, 2013; Dillenburg, Greene, & Erekson, 2003; Windolph, 2011). Windolph (2011) summarizes academic findings, noting that the transparency of SRs is characterized by "rare full disclosure of methodology, criteria, threshold values, etc." (p. 42). Representatives of non-governmental institutions have even called the industry of socially responsible investments "ridiculous" (Hawken, 2004, p. 3) due to the lack of standards, definitions, and specific regulation. Hawken (2004) criticizes a lack of transparency by fund managers regarding the screening and selection carried out by SRs agencies. His analysis finds that "over ninety percent of Fortune 500 companies are included in [Socially Responsible Investment] portfolios" (p. 18). As the individual portfolios hold only small parts of the Fortune 500 companies, the meaning of SRs must differ.

Of the authors that have criticized the lack of transparency in the past ten years, only a few put forward concrete elements about which an SR agency should be transparent. Most criticism remains vague or focuses on some indicators that an SR takes account of. This is the case despite the fact that the construction of an SR consists of identifiable methods and decisions. The result of an SR is the grade or the rank that rates every company. This result needs to be interpreted in order to understand what it stands for. Differences may include: Does it measure a relative or absolute concept of sustainability? Does it include actual environmental impacts? How does it define social responsibility? What is the influence of individual indicators on the result? How demanding is it about individual indicators? Is it a measure of strong or weak sustainability? Does is exclude some industries? All of these aspects depend on the construction of the SR.

In summary, sustainability is a global objective, but it is ill defined. The financial market is an effective lever to incentivize sustainable development. It can integrate sustainability based on SRs. To know what this integration actually incentivizes, SRs need to be transparent. Being transparent is an imprecise claim that needs to be substantiated. This thesis establishes a set of transparency criteria to identify whether SR agencies are sufficiently transparent or not. Sufficient transparency enables the interpretation of an SR. I analyze the composite indicator

framework to identify and structure methods and decisions in the construction of an SR with an influence on the results. I use these findings to set up transparency criteria for the construction of SRs. In a brief second step, the application of this set of criteria is demonstrated by reviewing five SR agencies, which have been certified by a quality standard for SRs.

1.2 Research Questions

There are three corresponding research questions which guide this thesis. Aiming at establishing transparency criteria, I ask:

(RQ1) *What information is necessary to interpret a one-dimensional measure that represents a multi-dimensional phenomenon like corporate sustainability?*

To answer this initial question, we need to know which elements influence the results of a composite indicator. Consequently, I investigate the underlying question first:

(RQ2) *How is a one-dimensional measure, which represents a multi-dimensional phenomenon like corporate sustainability, constructed?*

To apply the transparency criteria identified by answering (RQ1), the empirical part examines:

(RQ3) *Do SR agencies publish sufficient information for the interpretation of the SR results?*

These research questions correspond to one chapter each. I answer (RQ2) first, followed by (RQ1) and ending with (RQ3). The following section describes the approach and structure in more detail.

1.3 Approach and Structure

The three parts constituted by the research questions are complemented by a fourth introductory part. First, by extracting characteristics and the problem setting of SRs, I show that the composite indicator framework is an applicable research framework. Second, I identify steps in the construction with a direct influence on the results of composite indicators in general, and explain how a one-dimensional measure for corporate sustainability is constructed to answer (RQ2). Third, based on steps and decisions during the construction, I formulate transparency criteria of SRs to answer (RQ1). Fourth, I demonstrate the application of my criteria in an empirical review of SRs to answer (RQ3) for five SR agencies.

The structure of this thesis represents the four-fold approach. The composite indicator framework and the SR approach are compared in chapter 2. The SR approach is introduced in section 2.1 and the composite indicator framework is introduced in section 2.2. Section 2.3 demonstrates the equivalence of their constituents and their problem settings. An application of the composite indicator framework to SRs has not been carried out before. In order to analyze SRs, it is necessary to have a research framework that I can rely on to identify crucial aspects of an SR's construction and to use findings from this field of research.

Chapter 3 identifies and presents crucial steps in the construction of composite indicators to explain how a single measure for corporate sustainability is constructed. Five crucial steps are identified to be relevant for the interpretation of SRs (section 3.1). These are the theoretical framework[1] (section 3.2), the data selection (section 3.3), the derivation of weights (section 3.4), the normalization (section 3.5), and the aggregation of indicators (section 3.6). At the end of this chapter, I will have shown how a single measure – which represents a multi-dimensional phenomenon like corporate sustainability – is constructed, corresponding to (RQ2). A major part of the used literature analyzes assessments of sustainable development concerning nations. Its findings are applicable to equivalent assessments of companies. Based on its applicability, I use insights from the composite indicator literature to find out which methods are used and which decisions are made during the construction of an SR. This literature is complemented with insights from ESG-literature, statistical offices, and lifecycle analysis where appropriate.

Once the construction of SRs has been analyzed, the identification of information necessary to interpret SRs is a simple transfer. These elements of information form the transparency criteria. In a formal exercise, they will be reasoned (section 4.1). Section 4.2 presents the transparency criteria and section 4.3 justifies the criteria by anticipating two kinds of criticism.

In chapter 5, the application of the transparency criteria is demonstrated empirically. For the review, I apply the transparency criteria to five SR agencies that have voluntarily been certified by a standard called Arista 3.0 in order to counter criticism concerning their quality and transparency. This "voluntary quality standard for responsible investment research" is a concrete effort by the SR industry to "stimulate transparency" among SR agencies (Association

[1] Please note that 'theoretical framework' refers the first step in the construction of a composite indicator. The 'composite

for Responsible Investment Services (Arise), 2012a, p. 3). The application of my transparency criteria identifies (1) whether the standard ensures enough transparency to enable the interpretation of SRs and (2) where improvements are needed. The certification process requires the publishing of a so-called transparency matrix, where agencies answer questions about their SRs. This suggests that crucial parts of the construction are published. However, the empirical analysis finds that the transparency matrices do not enable the interpretation of the SRs. The empirical part is presented as a separate part of the thesis, thus a hypothesis (section 5.1) and a literature review (section 5.2) is presented in chapter 5. I introduce Arista 3.0 and the data source in section 5.3. After a presentation of the results (section 5.4), I discuss them in depth (section 5.5).

Many other interesting questions in the context of SRs are explicitly excluded from this thesis. It is tempting to evaluate the quality of SRs and whether their scope of ESG issues is sufficient. One could create a 'correct' SR, which then is compared to existing SRs in order to identify deficiencies. To do so would be an elaborate endeavor – one that has been made by SRs in the past two decades. It would also be a subjective analysis, given that there is no universally accepted definition of corporate sustainability.

Without setting up one's own SR, one may compare the concepts of corporate sustainability represented by different SRs. I found that there is insufficient published information to do so in a satisfactory way. Thus, I refrain from speculating about the quality of SRs and their interpretation. The interpretation of an SR requires information about its construction, i.e. a procedure encompassing various methods and decisions. I analyze which information is published without analyzing the content released by this transparency. In other words, I do not evaluate the construction of SRs, but rather the availability of information concerning the construction.

I also excluded the analysis of SR agencies at an institutional level. An institutional analysis looks at the ownership of an agency, its affiliations, and incentives created by its business model. This could identify causes for more or less transparency of an SR's construction. Again, my analysis is restricted to the transparency concerning the construction. I avoid speculating about the causes for more or less transparency. My empirical conclusions compare the transparency criteria to the information published by SRs. The conclusions can only state whether an agency's transparency is sufficient to interpret its SR.

1.4 Axioms

Basing the theoretical part on two axioms avoids a lengthy justification. The first one is:

Axiom 1: *It is possible to construct a one-dimensional measure for the multi-dimensional phenomenon 'corporate sustainability'.*

The axiom states the current practice of SRs. This practice is analyzed. By basing my analysis on this first axiom, I exclude researching whether corporate sustainability can be measured at all. The acceptance of the first axiom allows me to focus on the methods and decisions made during the construction. The goal is to determine, whether the transparency of SR agencies is sufficient regarding the construction of a one-dimensional measure, not if this can be calculated at all. Other thought-terminating clichés, e.g. a company can principally not be sustainable, are also excluded. This depends on the definition of sustainability, which may be a matter of degree rather than a binary relation.

Axiom 2: *A SR agency has the obligation to be sufficiently transparent so that it can be interpreted.*

The second axiom states the need for transparency. This axiom is comprehensible from different perspectives. Concepts like the principal agent theory demonstrate the need for transparency and the cost of information asymmetries. Looking at an SR's stakeholders, several interest groups appear to rely on its transparency. Customers need to be able to interpret the measure to use it in their financial analyses. Rated companies want to have a chance to improve their SR. Other stakeholders, such as the civil society, want to understand the incentives for companies caused by SRs. All stakes depend on information about the construction of SRs. More concretely, they require information about those parts of the construction, which influence the results, as this information is needed for its interpretation.

Similar reasons have lead an expert group to the inclusion of transparency in the Bellagio principles. These are ten principles for the assessment of sustainable development. They were agreed upon when the International Institute for Sustainable Development and the Rockefeller Foundations invited experts to Bellagio, Italy, in 1996. Instead of discussing a set of ideal sustainability indicators, they agreed on the Bellagio Principles. Principle 6 "Openness" and principle 7 "Effective Communication" stretch the need for transparency in the same way the second axiom does (Hardi & Zdan, 1997).

2 Sustainability Ratings and the Composite Indicator Framework

This chapter demonstrates the applicability of the composite indicator framework for the analysis of SRs. The search for a research framework was necessary, as analyses of the construction of SRs in the literature are not based on exhaustive scientific methodology. Thus, I looked for a framework that shared key characteristics of SRs. The research framework then enables a structured perspective on the construction of SRs. I can rely on corresponding publications to identify crucial aspects of an SR's construction and to use findings from this field of research.

First, I introduce the SR approach (section 2.1), including the agencies that create them, the data they are based on, and their key characteristics. Next, I present the composite indicator framework (section 2.2), including its constituents, the relationship between the variables it considers, and the resulting measure. Finally, I demonstrate the applicability of the composite indicator framework to the analysis of SRs (section 2.3) by assigning constituents to their equivalents and presenting a formal problem setting that applies equivalently to SRs and composite indicators.

2.1 The Sustainability Rating Approach

SRs analyze characteristics of companies that SR agencies consider relevant concerning corporate sustainability. A SR's outcome is a one-dimensional measure that represents corporate sustainability. Therefore, the following definition of an SR is used by this thesis:

> *Sustainability rating. A sustainability rating is a one-dimensional measure that represents corporate sustainability.*

Corporate sustainability is not defined precisely. This means that due to the lack of the 'correct' measure of corporate sustainability, almost any one-dimensional measure that is supposed to measure corporate sustainability is in fact a measure of corporate sustainability. It suffices to acknowledge that corporate sustainability aligns a company's actions with the notion of sustainability. However, only three aspects of this notion are generally accepted. As stated in the introduction, it is anthropocentric, multi-dimensional, and long-term oriented.

SRs are constructed and published by SR agencies. They use so-called environmental, social, and governance (ESG) data. I first describe SR agencies and introduce ESG data afterwards. I conclude this section by extracting key characteristics of any SR.

2.1.1 Sustainability rating agencies at financial markets

A SR agency is the creator of an SR. Overviews list between 15-30 such agencies on the European market (Döpfner & Schneider, 2012; Eurosif, 2014; Novethic, 2013; Schäfer et al., 2006) and more than 100 agencies globally (Sadowski et al., 2010). At the financial markets, the SR agency has the role of an information and research provider. In broad terms, they collect data, offer access to their databases, and compile SRs. Some of the SR agencies also offer consulting services and act as fund managers. This thesis analyzes the SR part of these agencies as defined above.

Traditional financial ratings are typically paid for by the rated company, which attracted criticism following the financial crisis in 2008. In contrast, the customers of SRs pay for access (Sadowski et al., 2010). Customers like fund managers, banks, and consultancies use an SR to evaluate companies. It may be input to a traditional financial analysis or it may determine financial products directly, as is the case for index funds that track a sustainability index. In both cases, the SR influences investment decisions.

The financial performance of investment strategies based on sustainability has been intensely investigated. One meta study by Deutsche Bank Climate Change Advisors (2012) finds that companies with high ratings for sustainability factors "are correlated with superior risk-adjusted returns" (p. 5). Gladman (2011) refers to several other meta-studies indicating that "on average, responsible investment methods perform on par with conventional techniques, neither outperforming or underperforming them on a regular and reliable basis" (p. 1). A recent meta-study by Clark et al. (2014) finds that "80% of the studies show that stock price performance of companies is positively influenced by good sustainability practices" (p. 7). The different findings about the relationship between sustainability and financial performance may depend on the definition of corporate sustainability.

Some components of corporate sustainability are plausible indicators for a financially successful company. An analysis of a workshop series with financial analysts by the UN Principles for Responsible Investment gives the following examples (UNPRI, 2013). In the environmental sphere, less resource consumption leads to fewer expenses. It also reduces regulatory risks

and may be seen as an indicator of a well-governed company. In the social sphere, good working conditions relate to productivity, innovative suggestions, and loyalty. Further, a balanced gender ratio indicates an effective use of the available talent pool. Lastly, governance measures like compensation schemes align management and long-term interests of shareholders.

Just like traditional financial analysts, different SR agencies consider different dimensions. Schäfer et al. (2006) differentiate between economically oriented concepts and normatively oriented concepts. The former see a business case for sustainability and measure a concept of sustainability, which serves as an indicator of companies that improve the classic triad of risk, liquidity, and return by reducing risks and increasing returns. Normatively oriented approaches see sustainability as a value in itself and measure a concept of sustainability based on ethical considerations.

In practice, this bi-polar differentiation has many facets. On the one side, normatively oriented agencies claim to improve risk-return profiles as well. On the other side, economically oriented agencies advertise the positive societal impact of their valuations. However, there are clear differences concerning the primitive maxim upon which the various agencies act.

In conclusion, SR agencies have a similar role to traditional rating agencies. There are numerous approaches to assess a company's sustainability. The interpretation of sustainability can aim to maximize profits, but may also prioritize normative motivations. The corresponding maxim influences the considered data. However, the population of data is generally referred to as ESG data.

2.1.2 Data and dimensions

ESG data originally referred to measures for environmental, social and governance issues. At the same time, ESG has become an independent term for sustainability aspects of many different kinds. It lacks a clear definition. Whether a measure is considered 'ESG' is in the eye of the beholder. The notion of ESG data encompasses measures that reflect sustainability risks and opportunities as well as responsible behavior (Bassen & Kovács, 2008).

ESG data is partially published by companies themselves, i.e. self-reported, and partially collected by SR agencies through questionnaires and external sources, e.g. NGOs and media. Coordinating entities set standards that aim to align reporting across companies. These entities

are typically investor-backed institutions, similar to reporting standard setters in financial accounting. One such entity, the Global Reporting Initiative (GRI) (2013c), published the fourth version of its reporting guidelines. The Sustainability Accounting Standards Board (2013) recently started to publish industry-specific reporting guidelines, such as for the pharmaceutical industry. The same approach is followed by the European Federation of Financial Analysts and the Society of Investment Professionals in Germany (2010). The latter two standards prescribe on which measures the companies have to report, while the GRI only requests reports on measures that the companies consider material for themselves. The standard with the lowest reporting requirements comes from the International Integrated Reporting Council (2013) in the form of a purely principle-based approach without fixed dimensions.

The GRI guidelines are the most established ones among the existing guidelines. GRI's newest guidelines 'G4' comprise 155 different measures as standard disclosure with more topics and indicators proposed in the GRI sector disclosures (2013c). Four of these measures shall illustrate the range of ESG-indicators:

- % of total employees covered by collective bargaining agreements (G4-11)
- Stakeholder management arrangements (G4-41)
- Direct greenhouse gas emissions in t CO_2e (G4-EN15)
- % of new suppliers that were screened using human rights criteria (G4-HR10)

These measures illustrate, that both quantitative and qualitative measures are part of ESG data. In order to align measurement processes, some quantitative measures build upon specialized accounting standards, e.g. the greenhouse gas protocol (World Business Council for Sustainable Development (WBCSD) & World Resources Institute (WRI), 2004). In other cases, it is up to the company to define and explain the measures it discloses, e.g. the definition of human rights criteria for its supplier screening.

2.1.3 Key characteristics of sustainability ratings

In order to find an appropriate research framework to analyze the construction of sustainability indicators, I extract key characteristics of the problem setting faced by SRs. The research framework should reflect all these characteristics. There are three characteristics. These are the purpose of measurement, the multi-dimensionality and fuzziness of corporate sustainability, and the aggregation into one dimension.

The first characteristic refers to the very basic purpose of SRs: Measurement is a process of association between an attribute of an object of analysis and a category (Hardegree, 2001). The category may be a group, a grade, a number, etc. Depending on the interpretation of the category, we differentiate between nominal, ordinal, interval, and ratio level of measurement[2] (Stevens, 1946). In the case of SRs, several attributes are measured on different levels of measurement. Eventually one measure is concluded, typically on the interval level of measurement.

The second characteristic of SRs is that they measure something that is not defined precisely and influenced by multiple dimensions. All SRs are measures of corporate sustainability but they interpret corporate sustainability differently. The research framework needs to be able to analyze the construction of SRs independent of the individual concept of sustainability in order to be applicable to various SRs. The notion of sustainability is multi-dimensional, anthropocentric, and with a long-term orientation (Moldan et al., 2012). It is fuzzy, because uncertainty is not only linked to data and occurrence of the event, but to the "event itself, which cannot be described unambiguously" (Munda, 2005, p. 978). This is at odds with dichotomous and unambiguous measurement approaches from traditional measurement theory. Dimensions of sustainability typically include but are not limited to the environmental, social, and economic sphere. Corporate sustainability implies that a company behaves in line with the global goal of sustainability. The data used by an SR usually includes environmental, social, and governance dimensions. The economic dimension may be part of it, but it is often left out to avoid overlaps with traditional financial analysis. The governance dimension includes prerequisites of corporate sustainability, e.g. the responsibilities of executive board members.

Third, aggregation delineates SRs from mere ESG research. As investor activities melt down to binary decisions (i.e. invest or not invest), a single measure is necessary to enable comparisons. Without noting that company a_j is more sustainable than company a_{j+1}, one cannot decide to invest in a_j based on its sustainability. Thus, aggregation is necessary for making comparisons which is in turn necessary for decision-making and decision-making is necessary for applications in financial markets. Consequently, SRs need to be one-dimensional.

[2] Note that the level of measurement depends on the interpretation of coefficient. I presuppose knowledge about the theory of levels of measurement. It is summarized in Appendix 2. Level of measurement is a synonym for measurement scale. I use the former expression to avoid confusion with scaling-transformations.

In conclusion, the three characteristics point towards a procedure with various parts. The purpose of measurement of something multi-dimensional and fuzzy requires defining a concept and identifying representative measures. In order to arrive at a one-dimensional measure, these measures will need to be weighted and aggregated. Any framework with an equivalent problem setting is likely to structure the construction process in a way applicable to the analysis of SRs. It may also use applicable methods and identify pitfalls and ambiguous decisions in the construction process.

2.2 The Composite Indicator Framework

A composite indicator is an aggregate of dimensions represented by indicators, and weights "on the basis of an underlying model" (OECD, 2008, p. 13). The term *composite index* is used synonymously in the literature. Composite indicators are typically calculated for nations to measure various phenomena, e.g. human development, technology achievement, scholastic achievement, competitiveness, or sustainable development (OECD, 2008).

Popular composite indicators are the Human Development Index, the Ecological Footprint, the OECD Better Life Index, the Ocean Health Index, and the Global Competitive Index (Hsu, Johnson, & Lloyd, 2013, p. 7). They all have in common, that several individual indicators are aggregated to form a composite that may assist decision-making (Parris & Kates, 2003). Due to the compensatory aggregation, composite indicators are typically measures of weak sustainability, even though they don't have to be compensatory (Munda, 2008, pp. 92ff.). The composite indicator framework is a collection of findings from social choice theory, measurement theory, and multi-criteria decision-making (Munda & Nardo, 2005; Munda, 2008). It also includes lessons from the aggregation of individual utility functions in social welfare economics (Ebert & Welsch, 2004). Multiple authors apply the composite indicator framework to analyze existing measures (Böhringer & Jochem, 2007; Booysen, 2002; Ebert & Welsch, 2004; Parris & Kates, 2003). The explicit discussion of composite indicators in the academic literature unfolded around the year 2000 when they were increasingly published to assist policy makers in various fields. A basic textbook does not exist. This shortcoming and the increasing interest in comparing nations based on composite indicators provoked the OECD (2008) to publish the *Handbook on Constructing Composite Indicators*, hereafter referred to as the OECD handbook.

The following subsections introduce constituents of any composite indicator, the relationship between variables, and the resulting measure.

2.2.1 Constituents

The main constituents of a composite indicator are the phenomenon, its dimensions, the variables, and the weights. I mostly use the terminology from the OECD handbook. To cope with different terminologies, I adapted direct citations with the corresponding terms using square brackets throughout the thesis. Page vii explains key terms to assist the reader further.

The *objects of analysis* refer to what the composite indicator is applied to. The literature typically discusses composite indicators that analyze nations. A different composite indicator score is calculated for each object of analysis. A scope should be defined to delineate objects of analysis.

The *phenomenon* refers to what is being measured by a composite indicator. A phenomenon is a concept of a notion. A notion's meaning is not universally agreed upon, but the general idea is known. Defining a specific concept is a prerequisite of the construction of a composite indicator. The latter steps are an explicit way of operationalizing a concept in order to measure it.

The phenomenon encompasses other, constituting phenomena. These constituents are *dimensions* of the phenomenon (OECD, 2008, p. 13). They should be collectively exhaustive with respect to the phenomenon (Bowker & Star, 2000, p. 10f.). Due to the multilevel characteristic of many phenomena, the dimensions can be considered as phenomena on their own with their own sub-dimensions. This implies that the process of construction is an iterative one. It needs to be carried out on each level of a phenomenon.

Once dimensions are defined on any given level, *variables* are assigned to each dimension. A variable may be a composite itself if several variables can be aggregated to represent a dimension. In general, the variable operationalizes each dimension. If a representative variable cannot be found, a proxy variable has to be identified.

Lastly, aggregation requires *weights* for each variable. Analysts derive them based on participatory or more sophisticated methods. The weights represent the significance of variables to the phenomenon. Weights can also be compared to each other. Thus, they need to reflect the relationship between variables. There are different types of relationships that a composite indicator can account for.

2.2.2 Relationships between variables

I encountered an inconsistent use of terms that describe the relationships between variables of a composite indicator. The terms used are comparability and commensurability. Their inconsistent use may be caused by the various settings, where relationships between variables and embodied values are described. Describing the use in philosophy, Hsieh (2008) notes that incommensurable values cannot be reduced to a common measure. Sen (1982, p. 203ff.) uses comparability to describe the relationship of individual utility functions in welfare economics. Ebert and Welsch (2004, p. 280) combine comparability and incommensurability by analyzing comparable, yet incommensurable variables that lack an "underlying scientific relationship". This is not consistent with the other definitions above. O'Neill (1993, p. 103) notes that the term incommensurability "has led to more confusion [...] than any other single word in the philosophical vocabulary". To avoid inconsistencies, I define incommensurability, incomparability, and their opposites. Although I thereby expose myself to (less important) terminological criticism, I avoid the more detrimental conceptual confusion. Afterwards, I describe the relationships between variables in composite indicators.

Incommensurability describes the lack of a common measure in principle. Incommensurability excludes the possibility of finding a causal relationship. *Commensurability* is the opposite of incommensurability. It describes the possibility of a causal relationship between variables with respect to a shared effect.

Incomparability describes the lack of a known relationship. This relationship may or may not be found and quantified in the future. *Comparability* is the opposite of incomparability. It describes the existence of a causal relationship between variables that has been quantified with respect to a shared effect.

These definitions allow three types of relationships between variables. They are treated differently in the composite indicator framework.

The first type of relationships is comparable, hence commensurable. Comparable variables are always commensurable and have a unit that refers to their influence on a shared effect. For example, the unit tons of CO_2-equivalents (CO_2e) aggregates greenhouse gases based on their climate warming potential over a period, usually 100 years. In the composite indicator framework, these variables are aggregated into one variable to represent one dimension.

The second type encompasses incomparable variables that are commensurable. In these cases, composite indicators try to estimate a causal relationship during the derivation of weights. The Ecological Footprint is such an example. In this case, there was no relationship between the environmental stresses of settlement, timber, food, seafood, and energy until the Ecological Footprint Network calculated the area that was needed for built-up land, forest, cropland, fishing, and to absorb CO_2 (Galli et al., 2008). Now, statements like 'energy production requires more area than settlements' are possible based on a shared effect and measure. Hence, these variables are rendered comparable based on the construction of the composite indicator.

The third type of relationship is incommensurable and hence incomparable. Due to the lack of a shared measure, no statement can be made whether a degree of freedom is better than/as good as/worse than a degree of equality. Clearly, the aggregation of these variables causes challenges and necessitates compromises compared to traditional measurements. The composite indicator framework offers a systematic approach to see where these compromises are made and how they are minimized.

2.2.3 Resulting measure

A composite indicator derives a single one-dimensional value for each object of analysis. It enables comparisons of the objects of analysis. The value may be measured on the ordinal, interval, or ratio level of measurement. It can be in the form of a final ranking, a grade, a number, or a percentage.

The aggregation of multiple variables into one measure is contested. Sharpe (2004) notes a "fundamental division in the indicators literature between those who choose to aggregate variables into a composite indicator and those who do not", with latter ones objecting "what they see as the arbitrary nature of the weighting process by which the variables are combined" (p.5).

As implied by Sharpe, the composite indicator framework is the framework of choice by the proponents of aggregation. Naturally, uncertainties increase with the level of aggregation. Skeptics further question the aggregation of incommensurable variables because it "inserts an ethical dimension that must be consistent with the stakeholder views" (Gasparatos, El-Haram, & Horner, 2009, p. 251). The composite indicator framework makes the necessary choices explicit and explores possibilities to reduce the ambiguity of the results. This thesis confers to

axiom 1 instead of entering the discussion of whether a one-dimensional measure can be derived at all.

2.3 Applicability of the Composite Indicator Framework

It is already noticeable that the composite indicator framework is a good fit for the analysis of SRs. Composite indicators share the three key characteristics of SRs. Their purpose is *measurement*, they measure *multi-dimensional and potentially fuzzy concepts*, and they aggregate indicators into *one measure*. To demonstrate the applicability of the composite indicator framework for the analysis of SRs, I assign constituents of SRs to their equivalents in the composite indicator framework. Afterwards a formal problem setting is presented, which applies equivalently to SRs and composite indicators.

2.3.1 Equivalence of constituents

The constituents of the composite indicator are the phenomenon, its dimensions, the variables, and weights.

The *objects of analysis* of SRs are companies. Their scope is delineated along the value chain.

The *phenomenon* of SRs is corporate sustainability. It encompasses multiple *dimensions*, e.g. the environmental and social sphere. It may have sub-dimensions, e.g. environmental performance, and sub-sub-dimensions like climate change impact. The composite indicator framework incorporates this multilevel characteristic by treating sub-dimensions as phenomena themselves. A *variable* is assigned to each dimension at the lowest level. Within financial markets, these variables are collectively referred to as ESG data. They can be quantitative, e.g. direct greenhouse gas emissions, and qualitative, e.g. a carbon risk assessment.

Weights need to be derived for SRs. Most variables from the ESG spheres are incommensurable. Hence participatory methods are usually applied.

One difference between typical SRs and the composite indicator remains. This is the use of exclusion criteria by SRs agencies. SRs sometimes exclude certain industries like the weapons industry. Such exclusion criteria constitute dictators in a lexicographic decision process (Munda, 2008, p. 4). They are not included in the composite indicator framework. They can

be seen as upstream filters and need to be discussed outside of the composite indicator framework.

2.3.2 Equivalence of the formal problem setting

A formal problem setting is the basis for a precise language. The following is based on Munda (2008, p. 7), who describes the problem setting for a composite indicator and relates it to multi-criteria decision-making. I complement it with the corresponding constituents of SRs. The differentiation into a global set of dimensions and a set of relevant dimensions is added to Munda's problem setting. This structures the derivation of relevant dimensions more clearly.

Composite indicators analyze objects. A is a finite set of N objects of analysis a_j with $j = 1, 2, \dots, N$. Objects of analysis correspond to alternatives in the multi-criteria decision-making framework. They are companies in the case of SRs.

Composite indicators measure phenomena. The phenomenon is a notion that is being defined as a concept during the construction of a composite indicator. In theory, the phenomenon's constituents are dimensions of the set G. G is a finite set of M dimensions m_i with $i = 1, 2, \dots, M$. Dimensions correspond to criteria in the multi-criteria decision-making framework. They are dimensions of corporate sustainability in the case of SRs.

In practice, a composite indicator measures a version of the phenomenon with dimensions of the subset G' with M' relevant dimensions $m_k, k = 1,2, \dots, M'$, with $G' \subseteq G$ and $M' \leq M$. The identification of this subset is influenced by the primary objective of the composite indicator. SRs differ in the set of relevant dimensions G' they consider.

Composite indicators are calculated based on variables. Each dimension is represented by one variable x. Variables correspond to the criterion score in the multi-criteria decision-making framework. They can be quantitative or qualitative ESG measures in the case of SRs.

Given the set A and set G' and assuming the existence of N objects of analysis and M' dimensions, it is possible to derive $N \times M'$ variables $x_{j,k}$ with $j = 1, 2, \dots, N;\ k = 1, 2, \dots, M'$. It represents the evaluation of the j-th company by means of the k-th dimension.

Some variables $x_{j,k}$ are transformed into scaled variables $\overline{x_{j,k}} = x_{j,k}/y_j$. The scaling variable y_j represents a characteristic of companies, e.g. sales.

Prior to aggregation, variables $x_{j,k}$ and scaled variables $\overline{x_{j,k}}$ are transformed into indicators[3] $I_{j,k}$. Company a_1 is better than company a_2 according to the k-th indicator if $I_{1,k} > I_{2,k}$.

A composite indicator CI_j is defined as the aggregate of all I_k for the j-th company. During this aggregation, each criterion m_k is weighted by a weight w_k, part of set H with $k = 1,2, \dots, M'$. Consequently, there are N results for the N objects of analysis.

Figure 1 shows the combination of set A and set G' via the variables $x_{j,k}$. The indicators of one company and the weights for each dimension form the company's composite indicator score.

Figure 1: Matrix presentation of a composite indicator.

	m_1	...	m_k	...	$m_{M'}$
a_1	$x_{1,1}$	...	$x_{1,k}$	...	$x_{1,M'}$
...	...	...	...	...	...
a_j	$x_{j,1}$	...	$x_{j,k}$	...	$x_{j,M'}$
...	...	...	...	...	...
a_N	$x_{N,1}$	...	$x_{N,k}$	...	$x_{N,M'}$

$$\rightarrow \begin{bmatrix} I_{1,1} & \dots & I_{1,k} & \dots & I_{1,M'} \\ \dots & \dots & \dots & \dots & \dots \\ I_{j,1} & \dots & I_{j,k} & \dots & I_{j,M'} \\ \dots & \dots & \dots & \dots & \dots \\ I_{N,1} & \dots & I_{N,k} & \dots & I_{N,M'} \end{bmatrix} \times \begin{bmatrix} w_1 \\ \dots \\ w_k \\ \dots \\ w'_M \end{bmatrix} \rightarrow \begin{bmatrix} CI_1 \\ \dots \\ CI_j \\ \dots \\ CI_N \end{bmatrix}$$

Source: (Munda, 2008). "×" represents any aggregation rule.

To sum things up, except for the exclusion criteria, all constituents of SRs have an equivalent constituent in the composite indicator framework. Furthermore, the problem settings are equivalent as demonstrated by the shared problem setting that corresponds to SRs and composite indicators. I conclude that the composite indicator framework is an applicable framework for the analysis of the construction of SRs. In fact, I treat SRs as composite indicators to structure their construction and learn about methodological choices.

2.4 Remarks

The composite indicator framework is a theory with a particular understanding of measurement. Colloquially speaking, composite indicators approach a phenomenon by defining its dimensions, assigning a variable to each dimension, and magically aggregating these variables into one measure. This contrasts with usual requirements of measurement problems. In fact,

[3] Following Ott (1978) I use the term variable for data before transformation and indicator after normalization step.

Arrow's impossibility theorem proves that it is not possible to derive a decisive result that fulfills basic measurement requirements when input data from different voters is measured at the ordinal level (Arrow, 1963). Sen (1982) extended this theory for cardinal data. In case of composite indicators, there are multiple dimensions instead of multiple voters. The findings remain valid, thus some basic measurement requirements are ignored when incommensurable variables are aggregated. The composite indicator framework acknowledges these findings (Ebert & Welsch, 2004) and offers a structured way of "measuring the immeasurable" (Böhringer & Jochem, 2007). Instead of denying the calculation of a measure for fuzzy notions such as sustainability, it points to the various decisions made by analysts to define their concept of sustainability and structures the construction of the corresponding measure. The decisions made by SR agencies influence the results and hence they need to be transparent. The identification of these elements is the motivation of this thesis.

Further, it should be noted that SRs accept a high level of approximation. Many variables subsume heterogeneous attributes because companies are extensive objects of analysis. Product lifecycle analysts perform a similar assessment for products. The level of approximation is a lot lower than the one of SRs. A trade-off between accuracy and costs exists and costs increase with the size of the object of analysis. The consideration of environmental performance can illustrates this point. It is defined as the aggregation of multiple environmental impacts caused by a company, and is inherently multi-dimensional. Environmental impacts include the consumption of resources like minerals/water/fossil fuels/etc. and the pollution of air/soil/water/etc. (UN et al., 2014, pp. 101ff.). Further breaking down the environmental impacts into sub-dimensions can allow for chemical, physical, and local differentiation. Interdependencies increase complexity further. For example, wastewater is dependent on the chemical properties of the residuals and their interdependent effects; resources can be differentiated into many sub-groups to account for different scarcities; water consumption is highly dependent on the local water availability. SRs constantly simplify in order to enable measurements at reasonable costs. I anticipated these difficulties by axiom 1, which states that measuring corporate sustainability is in principle possible nevertheless.

One last remark concerns readability. I mostly refrain from mentioning the need for differentiation of variables. Concerning the construction of composite indicators, I present a linear process without iterations. To do so, I need to ignore the multilevel characteristic of phenomena and the iterative nature of the process.

3 Construction of Composite Indicators

This chapter tackles the second research question:

(RQ2) *How is a one-dimensional measure, which represents a multi-dimensional phenomenon like corporate sustainability, constructed?*

I have shown in chapter 2 that the composite indicator framework is applicable to this problem setting. In this chapter, I present the construction of composite indicators in section 3.1 to identify five steps that have a direct influence on the results of SRs. These steps are the setup of a theoretical framework, the data selection, the derivation of weights, the normalization, and the aggregation of the constituents. These steps are discussed in sections 3.2 to 3.6.

Generally, I describe and discuss the steps by referring to the composite indicator literature. I transfer the steps to SRs and show relevant developments in this context. Examples mostly refer to corporate sustainability and parts thereof. Where appropriate, insights from ESG related literature, statistics, lifecycle analysis, and environmental economics complement my remarks. Above all else, it is the structure and methods from the composite indicator literature that guide this chapter. The research question is supposed to be answered by looking at composite indicators to find out how they construct a one-dimensional measure. This enables the identification of elements that have a direct influence on the results. The chapter is the basis to identify elements that need to be transparent to enable the interpretation of the SR.

3.1 Identification of Steps

The composite indicator literature discusses steps of the construction of a composite indicator. This section presents an overview of all steps discussed. Next, I identify those five steps that influence the results of composite indicators directly.

3.1.1 Steps in the literature

The literature distinguishes around ten steps in the construction of composite indicators. Depending on the focus of the publication, the suggested break down has a different emphasis. Table 1 roughly assigns steps from four publications to the OECD's break down. Steps with mere operational and organizational objectives are ignored. For example, Hsu et al. (2013) describe team-building measures during the construction of the Environmental Performance Index.

Table 1: Steps in the construction of composite indicators.

OECD, 2008	**Hsu et al., 2013**	**Zhou et al., 2006**	**Booysen, 2002**	**Böhringer & Jochem, 2007**
Theoretical framework *	Defining objectives & principles *	Defining environmental system *		
Data selection *	Designing a framework *	Variables selection & classification *	Selection *	
Imputation of missing data				
Multivariate analysis	Evaluating data quality			
Normalization *	Determining targets & calculating performance indicators *	Data collection and processing *	Scaling *	Normalization *
Weighting and aggregation *	Assigning weightings & aggregating an index *	Data aggregating *	Weighting and Aggregation *	Weighting and Aggregation *
Uncertainty and sensitivity analysis				
Back to the data			Validation	
Links to other indicators				
Visualization of the results				

Sources: as indicated. Own presentation. * marks steps that are relevant in this thesis. The terminology is taken over from the respective source without adapting it to this thesis' terminology. Steps are not congruent, thus the assignment is approximate.

The thesis focuses on steps marked by * in table 1. They are summarized into five steps: the theoretical framework, data selection, derivation of weights, normalization, and aggregation. Three of these steps correspond to the terminology of the OECD handbook. The derivation of weights is discussed separately from the aggregation to emphasize the importance of weights. As noted by Sharpe (2005), the "arbitrary nature of the weighting process" is contested. Thus, I discuss corresponding methods separately.

All of these affect the results of the construction exercise directly. Their order is not a matter of course. For example, normalization and aggregation influence the interpretation of weights. I chose to treat variables and weights first, because they are input data. Only after that, I discuss how this data is manipulated and aggregated by the normalization and aggregation.

The other steps are excluded for different reasons. The analysis of data and results should lead to adjustments of afore-mentioned steps. They do not influence the results directly and do not change their interpretation directly. SR agencies work with questionnaires and publicly available information. If a company does not report an aspect, it usually receives the worst grade. Thus, missing data is dealt with quickly and its discussion is not valuable. The last two steps of the OECD handbook present the results and compare them to other composite indicators. They are carried out once the results have been derived. Thus, these steps do not influence the results.

3.1.2 The five steps in this thesis

Each of the five steps identified above will be discussed in one of the sections 3.2 through 3.6. This section gives an overview of their content, their influence on the results, and their relationship to each other. The five steps are presented in table 2 along with their lead questions and an illustrative example for an SR. The example describes an environmental phenomenon because it can be presented without multilevel characteristics.

First, the theoretical framework describes the basis for all subsequent steps. It consists of the composite indicator's primary objective, the objects of analysis, and the phenomenon. These preliminary fixings are the basis for further decision-making in the construction of the composite indicator. Any composite indicator is based on the assumption that measuring the phenomenon for each object of analysis contributes to achieving the primary objective. For example in table 2, measuring the environmental performance of companies is assumed to predict financial performance. The phenomenon needs to be defined. A spelled-out definition can be complemented by a list of dimensions that may be relevant in theory. The primary objective, object of analysis, and phenomenon affect the results due to their influences on subsequent decisions. Metaphorically speaking, they span the room of the composite indicator construction.

Second, the data selection describes how to identify relevant dimensions and how to select corresponding variables. Relevant dimensions are the ones that are actually considered by the composite indicator. They are chosen based on a selection criterion that spells out the primary objective. Choosing variables operationalizes the phenomenon and determines which variables will eventually enter the calculation. To account for differences between the objects of analysis, a scaling variable is also chosen for some of the variables. As variables and scaled

Table 2: Five steps of the construction of a composite indicator as discussed by this thesis.

Name and Description	**Lead questions**	**Illustrative example**
Theoretical Framework (3.2)		
Defining the primary objective.	- What is the primary objective of the SR?	*Primary objective*: Predicting financial performance.
Defining the objects of analysis.	- Which objects of analysis are rated?	*Set of objects of analysis* A includes N companies a_j.
Defining the phenomenon.	- What is the definition of the phenomenon? - Which set of dimensions defines the phenomenon in theory?	*Phenomenon*: Corporate environmental performance. *Global set of dimensions* G includes M environmental domains m_i.
Data selection (3.3)		
Deriving relevant dimensions.	- Which criterion is the basis for the derivation of relevant dimensions? - Which set of dimensions defines the phenomenon in practice?	Selection criterion: Financial materiality. *Set of relevant dimensions* G' with M' *dimensions* $\mathrm{m_k}$. Dimension $\mathrm{m_1}$ is climate impact.
Selecting variables for each dimension.	- Which variables represent each dimension? - Which variables should be scaled? Which scaling variable should be used?	*Variable* $x_{1,j}$ is CO_2e emissions in tons. *Scaling variable* y_j is sales in € million. *Scaled variable* $\overline{x_{1,j}} = \frac{x_{1,j}\ t\ CO_2e}{y_j\ €}$
Derivation of Weights (3.4)		
Identifying weights for each dimension.	- Which weight is attributed to each dimension?	Equal weights for all dimensions. *Weighting coefficient* is $w_k = 1/M'$.
Normalization (3.5)		
Normalizing variables.	- Which normalization method should be used?	*Indicators* I_k are defined by the ranking method: $I_{k,j} = Rank(\overline{x_j})$
Aggregation (3.6)		
Aggregating indicators and weights.	- Which aggregation rule should be used?	Weighted arithmetic mean: *Composite Indicator* for j is $CI_j = \sum_{k=1}^{M'} w_k I_{k,j}$

Own presentation. Numbers in brackets refer to the section in this thesis. The example is illustrative.

variables are the origin of what is part of the aggregation rule, the data selection affects the results.

Third, the derivation of weights describes how to obtain the set of weights H. Weights represent a degree of significance with respect to the phenomenon. There are a variety of different methods leading to different weights. Weights influence the results because they are part of the aggregation rule.

Fourth, the normalization prepares the variables for aggregation. Different normalization methods exist and they lead to different indicator values. Normalization is sometimes referred to as standardization, although this is only one possible method of normalization according to the OECD handbook (2008). The indicator values influence the results, because they are part of the aggregation rule.

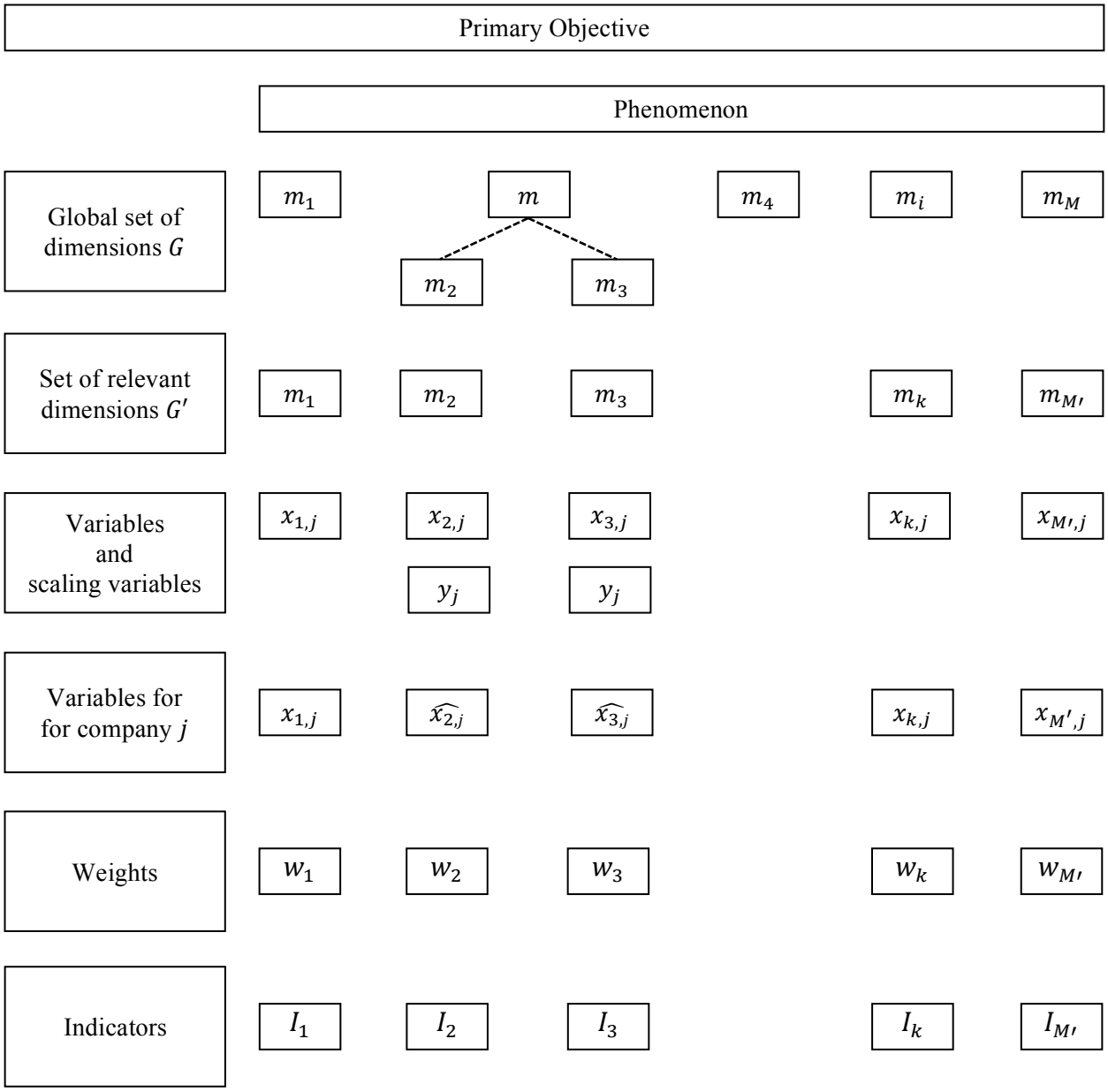
Finally, the aggregation of indicators and weights is based on an aggregation rule. Various aggregation rules exist and they imply different compensability among the indicators. Thus, the aggregation influences the results.

Figure 2 illustrates the relationship of the elements to each other. It shows the primary objective above all other elements, as it influences everything else. The phenomenon is defined by the global set of dimensions G. Most of these dimensions will have sub-dimensions. This multilevel aspect is illustrated by m_2 and m_3, which are both assigned to the non-numbered dimension m. The set of relevant dimensions G' does not necessarily include all dimensions from the theoretical definition, as illustrated by m_4. Consequently the number of dimensions decreases, i.e. $M > M'$.

All the relevant dimensions are assigned to a variable $x_{k,j}$. Some of the variables are also divided by a scaling variable, as illustrated by $\widehat{x_{2,j}}$ and $\widehat{x_{3,j}}$. Variables that already represent ratios (e.g. share of female top management) or that are on the ordinal level (e.g. other ranking results) remain non-scaled. This is illustrated by $x_{1,j}$, $x_{k,j}$, and $x_{M',j}$. The resulting set of variables for the composite indicator consists of scaled variables and non-scaled variables.

The dimensions' index number k counts all dimensions at the lowest level, i.e. the level that is used for the calculation of the composite indicator. For each relevant dimension, a weight w_k needs to be derived. Normalization transforms the variables into indicators I_k. Subsequently, there are M' relevant dimensions, variables, weights, and indicators.

Figure 2: Schematic presentation of a composite indicator.



Sources: (Hsu et al., 2013; Munda, 2008; OECD, 2008). Own presentation.

This section identified five steps in the construction of a composite indicator that have a direct influence on the results. Because of the framework's applicability for constructing an SR, I can now use insights from the composite indicator literature about these steps to find out which methods are used and which decisions are made during the construction of an SR. To this end, I follow the five-step structure, beginning with the theoretical framework.

3.2 Theoretical Framework

The theoretical framework is the basis for the subsequent construction. In motivating a theoretical framework as the starting point, the OECD handbook notes: “What is badly defined is likely to be badly measured” (OECD, 2008, p. 22). A theoretical framework is also proposed by the Bellagio principles, because “[f]irst, it helps determine priorities in the choice of indicators; and second, it triggers the identification of indicators which may be more important in the future” (Hardi & Zdan, 1997, p. 10). It is influenced by the audience (Parris & Kates, 2003, p. 22; Ramos & Caeiro, 2010, p. 158).

The theoretical framework consists of an explicit primary objective, the delineation of the objects of analysis, and the definition of the phenomenon. The primary objective describes the purpose and goal of the construction (Hsu et al., 2013, p. 10). The objects of analysis need to be consistently delineated so that the concept of a company is the same for all companies. And the phenomenon states *what* the composite indicator is supposed to measure (Hsu et al., 2013, p. 16; OECD, 2008, p. 22). Obviously, any composite indicator is based on the assumption that the measurement of the phenomenon for each object of analysis contributes to achieving the primary objective. SRs always measure a concept of corporate sustainability. However, this needs to be defined theoretically. The theoretical definition can be written-out or communicated by an explicit list of dimensions, the global set of dimensions G.

Note that these elements of a theoretical framework are not fixed. The ones discussed in this thesis are recommended by the developers of the Environmental Performance Index (Hsu et al., 2013) and the OECD handbook (2008). To first define the phenomenon with a global set of dimensions is my proposal to make the subsequent selection of relevant dimensions more explicit. The following sections discuss the three elements of a theoretical framework.

3.2.1 Primary objective

The primary objective is the main reason to construct and publish the results of a composite indicator. It is the reason customers use the results. Primary objectives of composite indicators can be measurement, communication and consensus building, issue framing, advocacy, decision making and management tools, analysis and research, and comparisons (Hsu et al., 2013, p. 14f.; Parris & Kates, 2003, p. 15).

The primary objective has an influence on *every single decision* in the construction of a composite indicator. When the primary objective is to construct a measurement tool, dimensions need to be measurable. When framing an issue, measurability is less critical. Instead, one might keep the number of dimensions low for understandability. Agenda setting and political objectives influence the construction of the composite indicator in support of the analysts' political goals. Here it becomes clear that the primary objective will not necessarily be spelled out completely. However, the audience of the composite indicator has to be enabled to decide whether the primary objective of the construction matches their own. In addition, an explicit primary objective is helpful internally for the consistent construction of a composite indicator (Hsu et al., 2013, p. 14).

SRs are mostly intended to assist in making investment decisions. Yet, this does not sufficiently define the primary objective as not all investors share the same objective. Schaltegger and Figge differentiate investment funds (cited in Koellner, Weber, Fenchel, & Scholz, 2005, p. 65). To name three: "Environmental technology funds" try to identify products addressing environmental problems, e.g. solar panel producers; "Eco-efficiency funds" invest in companies with the best eco-efficiency in their industry, which can be any industry; "Ethical funds" base their investment decisions on moral principles (Koellner et al., 2005, p. 64). This differentiation coincides with the differentiation into "economically orientated" and "normatively orientated" concepts described by Schäfer et al. (2006). To demonstrate the influence and importance of making the primary objective explicit, I consider two stylized examples throughout the thesis based on the latter differentiation.

In the first example, the customers of the SR, i.e. investors, want to maximize monetary returns. Therefore, the composite indicator needs to predict financial performance. For example, the SR agency Robecosam's *Corporate Sustainability Assessment* determines who is included in the Dow Jones Sustainability Index. This index is in turn the basis of various index funds. Robecosam claims that these financial products outperform other stock indices (2013, p. 9). This exemplary primary objective will be called *prediction of financial performance*.

In the second example, the customers consider social welfare. In economics, this approach is known as the social planner perspective. This group of investors intends to incentivize corporate responsibility for moral reasons. The SR then needs to measure companies' influence on social welfare. Social welfare is in turn defined by a concept of sustainability. Hence, it is affected by the corresponding concept of corporate sustainability, including the environmental

and social performance. For example, the SR of Oekom builds upon the Frankfurt-Hohenheim Guidelines, a list of guidelines for the assessment of corporate responsibility (Oekom, 2012). The authors of these guidelines aim to transform the economy into a system based on moral values and improving its service to humanity (Balz et al., 2000, p. 2ff.). In economic terms, the proposed measures aim to evaluate social welfare effects rather than predicting financial performance. Thus, I call this second exemplary primary objective *evaluation of welfare effects*.

3.2.2 Objects of analysis

Composite indicators are typically applied to nations (Hsu et al., 2013; Munda, 2008). Similar approaches analyze products, processes, individuals, projects, policies, and companies. For example, product lifecycle analysis concerns market- and transformation-activities during the lifecycle of a product (Baumann & Tillman, 2004). Footprints can be calculated for individuals, regions, and nations (Galli et al., 2008).

Two decisions need to be made about the objects of analysis. First, which set A with N objects of analysis a_j is considered by the SR? Second, how is each individual object of analysis delineated? Both are shortly explained by the following paragraphs.

First, SRs evaluate companies. The set A defines the sample of N companies a_j included in the SR. If an SR includes market-listed companies only, one may refer to the included indices. In addition, most SRs differentiate between industries. This limits the differences between the objects of analysis within one SR and makes comparisons easier. The population of companies under consideration is relevant for the construction of an SR because different industries and company sizes may influence the evaluation criteria. In the composite indicator theory, the consideration of distinct industries leads to different samples and implies that the agency constructs one SR for each sample. These SRs may be similar in some parts and different with respect to industry-specific dimensions.

Second, the individual object of analysis is defined by its scope in order to delineate it. The delineation of a company is typically based on a function-oriented definition. Instead of region-oriented definitions, function-oriented definitions take all processes into account "which are related to a given function or social demand, both upstream and downstream with respect to a core process which is directly fulfilling this function" (Wrisberg & de Haes, 2002, p. 59).

The scope of a company is defined correspondingly along the value chain. A SR can cover several possible scopes. One extreme is the inclusion of all activities linked to the company, from the extraction of the resources to the disposal of the product. This concept is known by product lifecycle analyses. The so-called *cradle-to-grave approach* avoids problem shifting due to outsourcing of resource-/emission-intensive processes, but requires extensive data (Hermann, Kroeze, & Jawjit, 2007, p. 1787). The other extreme is to restrict the scope to the (legal) entity. Criteria to delineate an entity's boundaries are widely discussed in financial accounting, e.g. the accrual principle and the consolidation. Environmental accounting standards such as *The Greenhouse Gas Protocol* refer to approaches from financial accounting, namely the control approach and the equity share approach (WBCSD & WRI, 2004, pp. 16ff.). According to the control-approach, the scope of a company encompasses entities controlled by the company. The equity share approach builds on the ownership principle and accounts proportionally for entities that the company owns share of.

Ultimately any scope may delineate the object of analysis. Regarding SRs, it seems advisable to include the supply chain. Otherwise, comparisons of companies are difficult due to different degrees of vertical integration and subsequent environmental impacts in the supply chain. There also seems to be agreement, that the supply chain is part of a company's responsibility regarding human rights (United Nations Global Compact & Business for Social Responsibility, 2012). Hence, a company should manage its supply chain sustainably; otherwise, it may face reputation and other risks.

Both decisions may influence the SR's resulting ranking. A broader scope can favor or penalize companies depending on their suppliers. The consideration of a certain sample will influence the choice of relevant dimensions.

3.2.3 Phenomenon

The "phenomenon and its sub-components" (OECD, 2008, p. 22) are at the core of the theoretical framework. It clarifies, what the composite indicator is supposed to measure. The definition of the phenomenon is independent of the primary objective. However, in principle the construction of a composite indicator relies on the assumption that measuring the phenomenon contributes to achieving the primary objective.

The OECD handbook considers the definition of the phenomenon crucial for the composite indicator's relevance, credibility, and interpretability (2008, p. 49). The phenomenon should

be specified by “key components” (Hsu et al., 2013) or “dimensions” (Munda, 2008; OECD, 2008). The explicit definition of the phenomenon and its constituents “should give the reader a clear sense of what is being measured by the composite indicator” (OECD, 2008, p. 22). Hsu et al. (2013) note that the definition of “core categories” helps the audience to understand the meaning of a composite indicator (p. 34). This understanding is not trivial, particularly in the context of sustainability. Parris & Kates (2003) find that “the ambiguity of sustainable development” and “the plurality of purpose in characterizing and measuring sustainable development” hinder the definition of a universally accepted indicator set in the context of sustainability (p. 23). They stress that it is all the more important how the phenomenon is defined.

The following section first demonstrates differences of spelled out phenomena based on an illustrative example for SRs. Further, I argue that it is advantageous to define all possible dimensions in theory before some of these dimensions are selected in the next step.

For this thesis, I broadly defined that corporate sustainability is corporate behavior that is in line with the notion of sustainability. For the construction of an SR, this needs to be specified. Corporate sustainability may be defined as a relative concept, where the most sustainable company of a sample is considered sufficiently sustainable. This is called the best-in-class or relative best-in-class approach (Goodall, 2005). An absolute best-in-class approach uses a threshold that is independent of the sample. In this case, industry leaders are identified if they are above a threshold that is independent of the sample (Oekom, 2012).

Different definition can be illustrated by considering the sub-phenomenon environmental performance. The literature presents different definitions (Cho & Patten, 2013, p. 445; Ilinitch, Soderstrom, & E. Thomas, 1998; Tyteca, 1996; O. Weber, Koellner, Habegger, Steffensen, & Ohnemus, 2005). Academics usually define environmental performance as the aggregate of environmental impacts, which are consequences of a company’s interaction with the environment lasting beyond the reporting period. Delmas et al. (2013) note that some SRs define environmental performance by looking at the implemented processes and practices of companies with regards to certain dimensions. Thus, instead of looking at the amount of carbon emissions, these SRs look at the strategy and actions a company implements. Delmas et al. (2013) show empirically that this definition leads to a different ranking compared to the one based on actual environmental outcomes. Furthermore, there is a third group of SRs that analyze whether a company discloses figures on carbon emissions (Rose, van Ast, Bolster, Knight, &

Bennett, 2013). In this case, environmental performance is defined as a measure of disclosure and has nothing to do with the actual emissions.

A spelled out definition should be complemented by a theoretical list of dimensions that belong to this phenomenon. I refer to this as the global set of dimensions G. It is an exhaustive list of dimensions of the phenomenon. The SR will typically not take all of these dimensions into account. However, it is helpful to name the ones that come into question in order to reason the exclusion of the ones that are not relevant.

When setting up a set of dimensions, the delineation between the dimensions needs to be mutually exclusive and collectively exhausting (Bowker & Star, 2000, p. 10f.). Collectively exhausting means that the complete phenomenon is represented. Mutually exclusive means that there are no overlaps. A top-down definition uses a division criterion to ensure mutual exclusiveness. A bottom-up definition gathers relevant dimensions through a poll or brainstorming exercise. The bottom-up approach cannot exclude overlaps and deficiencies in principle. Assumptions about the expertise of the participating individuals need to suffice in order to make sure that dimensions are collectively exhausting. Mutual exclusivity is important, as all variables should increase the composite indicators information value. If the dimensions overlap, the double representation of some aspects at the cost of others is difficult to avoid. In turn, this can reduce the relative weight of other indicators during aggregation. Depending on the dimensions that are implicitly down- or upgrade, some companies may be penalized or rewarded. This can be avoided by finding mutually exclusive dimensions.

It may seem unrealistic to look for a global set of dimensions that exhausts the phenomenon independent of the primary objective. However, preliminary work has been carried out in many areas of sustainability. In the area of social sustainability, SRs typically use reference norms to find a global set of dimensions. Examples of reference norms are the *Universal Declaration on Human Rights*, the *UN Global Compact*, or the *OECD Guidelines for Multinational Enterprises*. These reference norms are the result of a long opinion-forming process including various stakeholders. Hence, all stakeholders may agree with these dimensions. Subsequently, they are useful to set up a global set of dimensions. If everybody agrees that the set of human rights is the one laid out by the *Universal Declaration on Human Rights*, then this is the basis for the selection of relevant dimensions later on. In the area of environmental sustainability, other actors have classified the environment. Examples include statistical offic-

es, sustainability reporting standard setters, and product lifecycle analysts. Their global sets of dimensions are presented in appendix 1.

To sum things up, the primary objective is the very beginning of every construction. This primary objective is supposed to be achieved by measuring the phenomenon for every object of analysis. Concerning the objects of analysis, set A describes the considered sample and the scope delineates the individual object of analysis. Lastly, the phenomenon and sub-phenomena need to be defined. To this end, it is helpful to agree on a global set of dimensions that may be taken into account theoretically.

3.3 Data Selection

After the theoretical framework has been established, composite indicators are constructed on this basis. This section is about data selection. It selects relevant dimensions and corresponding variables based on the definition of the phenomenon (OECD, 2008, p. 23). If the phenomenon was formally defined by a global set of dimensions, this selection process simply means that the relevant ones are chosen based on a selection criterion. Note that the data selection is a methodological procedure, which is separate from the data collection where data is actually collected.

3.3.1 Set of relevant dimensions

In order to select relevant dimensions, one needs to acknowledge the difference between a theoretical concept and the concept that is actually measured. In other words, instead of assessing every social issue and every possible interaction with the environment, an SR considers a subset only. Borucke et al. explicitly acknowledge this difference for the Ecological Footprint (2013, p. 519). It is due to the measurement expenditures for marginally contributing constituents of the phenomenon, the availability of data, and the primary objective. Consequently, the dimensions that are considered by an SR are not the same as the dimensions of the phenomenon's theoretical definition. I refer to the subset as the set of relevant dimensions G'. It includes a selection of M' dimensions from the M dimensions of set G.

The selection exercise of those dimensions that are relevant is guided by the primary objective. The literature suggests to define a selection criterion to spell out how the primary objective affects the selection of dimensions with respect to a phenomenon: "It should be as precise

as possible and should describe the phenomenon being measured" (OECD, 2008, p. 22). The composite indicator literature does not provide a method or best practice that helps to identify relevant dimensions. In fact this is impossible; such a method would have to be an omniscient tool that identifies the 'correct' dimensions for different phenomena. Therefore, the construction of an SR has to rely on the selection criterion and subsequent case-by-case decisions. The following example illustrates defining a selection criterion for SRs.

The example starts out with materiality. Materiality is a generic decision-criterion. It is fulfilled if a dimension affects the primary objective by a degree sufficiently high to justify its inclusion. In financial accounting for example, audits are carried out for material figures only. This is defined by a threshold percentage of the overall audit, e.g. 1% of sales. In other words, materiality tests if a dimension is sufficiently relevant to be included in the SR.

The GRI reporting guidelines emphasize materiality to increase the relevance of corporate sustainability reports (GRI, 2013c, pp. 3, 8). It asks companies to report material aspects only. Material aspects in GRI are identified with respect to "the organization's significant economic, environmental and social impact; or […] the assessments and decisions of stakeholder" (p. 17). This definition addresses the divergent interests of stakeholders of corporate sustainability reporting (GRI, 2013b, pp. 35–37). A SR cannot serve divergent objectives, as this usually leads to different sets of dimensions. The participants of a workshop series by the UN Principles of Responsible Investment conclude that "the identification of material issues remains an art rather than a science" (UNPRI, 2013, p. 6).

Materiality is always dependent on something else. In the composite indicator framework, this is the primary objective[4]. The influence of the two exemplary primary objectives can be made clear by the concept of external effects. Negative external effects constitute the discrepancies between *prediction of financial performance* and the *evaluation of welfare effects*. I the absence of external effects, welfare effects were internalized and would affect financial performance directly. For example, the emission of greenhouse gases should be included to predict financial performance, if they were internalized by regulative measures. Without a risk of in-

[4] Materiality might also depend on the objects of analysis, e.g. if companies of different industries are considered. I assume one group of companies.

ternalization, the same SR excludes the dimension[5]. The decision criterion in this case is *financial materiality*. If, on the other hand, the primary objective is the *evaluation of welfare effects*, greenhouse gases are material irrespective of the legislation regarding internalization because they affect social welfare. The decision criterion in this case is *social materiality*.

To sum things up, relevant dimensions are selected based on a decision criterion. It spells out the primary objective and is applied to the theoretical definition of the phenomenon. Even if the decision criterion is precise, it can only by a guide for the identification of relevant dimensions. Ultimately, the selection procedure proceeds by case-by-case decisions. No method exists to identify relevant dimensions in a general way. At the same time, the selection of relevant dimensions decides about the inclusion or exclusion of sustainability aspects. The exclusion of a dimensions may favor or penalize a company. In turn, the selection of relevant dimensions has a direct effect on the SR score.

3.3.2 Variables and scaling variables

Once the set of relevant dimensions G' has been identified, the identification of variables x_i is dependent on technical considerations. A variable is "a constructed measure stemming from a process that represents, at a given point in space and time, a shared perception of a real-world state of affairs consistent with a given [dimension]" (Munda, 2008, p. 6). Composite indicators can use quantitative or qualitative data as a variable (OECD, 2008, p. 23).

Each dimension is operationalized by one variable. If the dimension is very specific, e.g. energy consumption, the identification of a variable is without problems. Broader dimensions without sub-dimensions require the identification of a single variable to represent the whole dimension. This variable is often a proxy (OECD, 2008, p. 23). For example, health and safety may be measured by sick days or accidents. They do not describe the health of workers nor the safety itself[6], yet they represent a proxy for this dimension.

---

[5] With the current state of climate change negotiations, it may even be necessary to differentiate between emissions in countries with a price on carbon and countries without a price on carbon. This is completely independent of the fact that greenhouse gases have the same effect on the climate independent of the location of emission.

[6] Sick days do not account for pain, minor chronic diseases, or the psychological conditions. Accidents account for worst cases, while safety describes the absence of risks.

If more than one variable is assigned to one dimension, there are two possibilities. If the variables are comparable based on an underlying scientific relationship, they are aggregated and treated as one variable (Ebert & Welsch, 2004). If the variables are not comparable, they are treated as two sub-dimensions. For example, the variable share of women in the workforce and the variable share of women in the executive board may be assigned to the gender diversity dimension. As they cannot be aggregated based on a scientific relationship, the two variables are considered as two sub-dimensions.

The objective of each variable indicate the desired development (Krajnc & Glavič, 2005, p. 553; Munda, 2008, p. 6). The objective seems trivial. Toxic waste, child labor, social inequalities, etc. always have a minimization objective ("the less the better"). Recycling, supplier audits, social equalities, etc. always have a maximization objective ("the more the better"). All objectives are taken into account during normalization, so that indicators entering the aggregation rule all have the same objective, usually a maximization objective. However, this changes if variables are context-dependent. Use of fertilizers may be beneficial or not, dependent on the state of an ecosystem. To increase complexity, impacts may be interdependent (OECD, 2008, p. 22). The analysts have to be certain about the objective of each variable, otherwise its inclusion does not contribute to the composite indicator.

It is challenging to find a variable for each dimension. Additionally, there are concerns about the quality of the data. The discrepancy between available and ideal data is omnipresent. Esty et al. (2005) summarize properties of ideal data sets in the environmental sphere, "These indicators would include all relevant aspects of functioning environmental systems, be distinct in their cause-effect relationships, permit aggregation, reflect the diversity of circumstances across political jurisdictions […], be easily quantifiable, and scale-neutral" (p. 12). On a related note, Munda (2008) demands "precise, certain, exhaustive and unequivocal" data (p. 64).

In the absence of an ideal data set, quality characteristics help to evaluate the data. The OECD handbook proposes six quality characteristics (OECD, 2008, p. 46ff.):

- *Relevance* assesses the value of the variable and its descriptive power with respect to the dimension.
- *Accuracy* is described by different kinds of errors in the data and is also endangered by subjective data sources.
- *Timeliness* reflects the time between availability and reporting period.
- *Accessibility* evaluates how much effort is needed to collect the data.

- *Interpretability* depicts the ease of understanding the data.
- *Coherence* means that data is "at least reconcilable" and may be hindered by changing definitions and methodologies over time or sources.

Each variable should be evaluated against all of these quality characteristics. In the context of SRs, coherence and accuracy are particularly critical. As data is often self-reported, companies may bias the data in their favor and thereby accuracy is endangered. Audits prevent these biases to some degree. Further, different measurement processes of the companies jeopardize coherence. Measurement processes need to define the same measures, interpret the effect in the same way, and relate to the same scope of the object of analysis. Reporting standards address these issues. For example, the greenhouse gas protocol defines the greenhouse gases, the effect of climate warming potential, and three different scopes (WBCSD & WRI, 2004). Unfortunately, few other reporting standards exist for the measurement of environmental data.

Even if variables were measured coherently, the differences of the objects of analysis remain. Scaling variables shall account for these. Scaling accounts for differences across objects of analysis, i.e. companies. It divides a variable $x_{k,j}$ by the factor of a scaling variable y_j. The resulting scaled variable is denoted by $\widehat{x_{k,j}}$.

Scaling variables are measures that are used to account for the dissimilarity of the objects of analysis', for example the different size of companies (Koellner et al., 2005, p. 61). The size of a company can be represented by financial parameters, the number of employees, working hours, or product quantities. In case more details should be accounted for, scaling variables may also account for the temperature at a factory's location, the kind of process, or the industry (e.g. Scalet, Garcia Muñoz, Sissa, Roudier, & Delgade Sancho, 2013). The selection of one scaling variable is "delicate" (Munda, 2008, p. 11). The following paragraphs discuss the selection of a scaling variable and its consequences.

Concerning environmental performance, differing degrees of vertical integration can change a variable's amplitude considerably. Consider car-producing companies A and B. A has outsourced the production of engines to India; B produces its own engines in France. Assume the production of engines causes half of the carbon emissions in the production of one car. To *evaluate the welfare effects* of these two companies, it does not suffice to measure the entities' internal emissions. Let us further assume that carbon emissions were to be charged with 70€ per ton within the European Union only. If the composite indicator aims to *predict financial performance*, the consideration of internal emissions suffices because there is no financial

burden on carbon emissions in India. B will have a financial disadvantage due to its carbon emissions within the EU. This financial disadvantage should not be scaled away. Thus dividing by sales will leave A better off and rightly so.

An alternative scaling variable for comparisons within one industry is the product quantity. Let us assume that Company A from above produces safe luxury vans carrying seven passengers while B produces low-cost, urban lightweight cars for two passengers. A has large margins while B struggles to generate any profit at all. Now if environmental impacts are scaled per car, B is favored because the production of a small car tends to cause less carbon emissions compared to a bigger car. Considering the car's carrying capacity as a scaling variable, A is favored because it only produces one engine, four wheels, etc. for seven passengers compared to a car that seats two passengers. Due to these difficulties, lifecycle analysts define a "functional unit" such as "packaging for 1000 liters of liquid" for scaling (Koellner et al., 2005, p. 61). With respect to cars, the functional unit is difficult to grasp, because people value properties like speed, comfort, reliability, and safety differently. Gasparatos (2010, p. 1616) concludes that an economic parameter used for scaling cannot be free of bias and ethical judgments.

To sum things up, the discussion about the correct scaling variable has not reached a conclusion albeit its long history (Schmidheiny, 1996). It is important to realize, that the choice of the scaling variable is not redundant and not neutral. A different scaling variable may change the resulting ranking of the composite indicator. Of course, the same applies to variables. Ultimately we are left with the same situation as for the identification of relevant dimensions. Even with quality characteristics assisting the selection of variables, it can only be used as guidance. Ultimately, the selection of variables remains a case-by-case decision. No method exists to select data in a general way. At the same time, the data selection has a decisive influence on the outcomes because it identifies the data that enters the calculation.

3.4 Derivation of Weights

The previous sections described how a theoretical framework builds the basis for the construction of an SR and how dimensions and variables are selected. The variables $x_{j,k}$ for all companies a_j and all dimensions m_k are one of two input arguments to the aggregation rule. The second one is the set H consisting of M' corresponding weights w_k. Consequently, there is one

weighting factor for every relevant dimension m_k and hence for each variable $x_{j,k}$. This section describes the derivation of set H. First, I introduce the different meanings of weights. Second, I present a three alternative methods that may be used to derive weights.

The OECD handbook (2008) notes that weights make the composite indicator's primary objective explicit, just like the identification of relevant dimensions. This indicates the influence of the primary objective once again. In addition, different methods lead to different weights. It is recommended to clearly state the way the weights are derived because weights are essentially value judgments, "no matter which method is used", thus denying the existence of objective composite indicators (p. 31f.).

Munda and Nardo (2005) questioned the meaning of weights in linear aggregation rules. They differentiate between the weights having the meaning of importance coefficients and weights having the meaning of trade-off ratios. If weights represent the concept of importance, they apply to the variable independent of the variable's value[7] (Munda & Nardo, 2005, p. 6). When weights have the meaning of importance, they are based on the variable's significance regarding the phenomenon. It does not take into account the relationship to other variables. In contrast, trade-off ratios are based on one-to-one comparisons. They consider the relationship of variables to each other and may even be scale-dependent, i.e. their value changes if the magnitude of the variable changes.

For example, in case of importance coefficients, the weight for climate change impact measured by CO_2e would be based on its significance to environmental performance, e.g. $w_{CO_2e} = .15$. Equally, the weight for wastewater would be based on its significance to environmental performance, e.g. $w_{ww} = .3$. The derivation of importance weights is based on the assumed relationship between a variable and the phenomenon. In contrast, the derivation of trade-off ratios is based on the assumed relationship between the variables. One would need to assess how much more important CO_2e is compared to wastewater. To confirm $.15$ and $.3$ as weights with the meaning of trade-off ratios, the one-to-one comparison needs to conclude that CO_2e is twice as important as wastewater. In order to derive trade-offs consistently for all

[7] One could think of importance coefficients as estimating relationships between dimensions instead of variables. This is not correct if proxy variables are used. Additionally it would be imprecise, because mathematically indicators are weighted, not dimensions.

variables of a phenomenon, methods need to assess all one-to-one comparisons of all variables (Munda & Nardo, 2005, p. 8, 2009). This is the only way to use compensatory aggregation rules without a theoretical inconsistency. It will be discussed further in section 3.6.

The composite indicator framework takes three relationships between variables into account. First, comparable variables are aggregated right away and treated like other variables. Second, variables that are commensurable but not comparable may be weighted based on monetary or biophysical relationships. Corresponding methods are applied by environmental economists (Pearce & Turner, 1990), popular composite indicators (Booysen, 2002; Borucke et al., 2013; Gasparatos, El-Haram, & Horner, 2008), and product lifecycle analysts (Goedkoop et al., 2013; Margni & Curran, 2012). In the context of SRs, these approaches have not been applied because quantifiable, causal relationships based on one measure describing 'corporate sustainability' do not exist. Therefore, I focus on methods for variables that are incommensurable, which may be weighted equally or based on participatory methods.

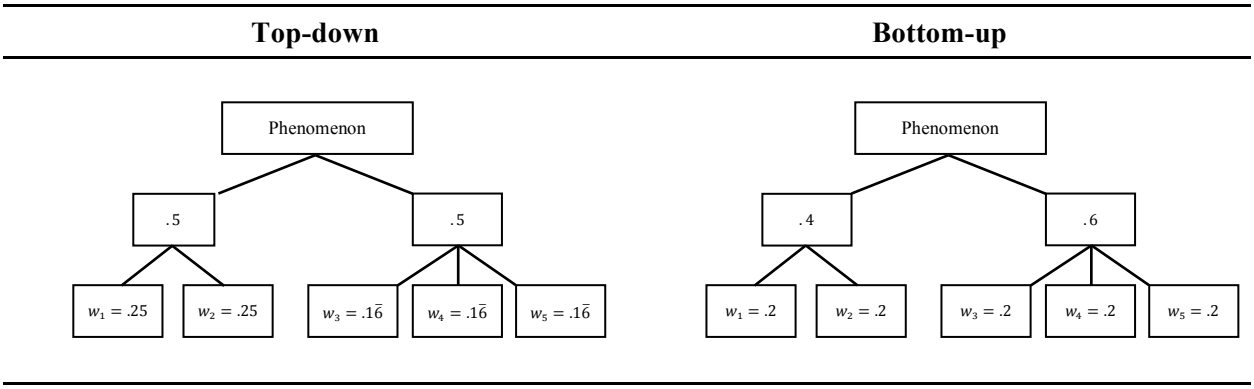
Variables and dimensions are incommensurable due to value pluralism. Examples of incommensurable dimensions are equality and fairness (Hsieh, 2008). Corresponding examples of incommensurable variables are the percentage of female board members and ratio of executive pay to worker pay. A causal relationship between these aspects does not exist. Furthermore, commensurable variables may be treated like incommensurable variables, if knowledge gaps hinder their aggregation based on causal relationships.

Three methods are presented below to describe how agencies may derive weights for incommensurable variables to construct SRs. These are the equal weighting method and two participatory methods, the budget allocation method and the analytical hierarchy process (AHP). I discuss the advantages and disadvantages of the methods and conclude that a superior methods also needs more information to be comprehended.

3.4.1 Equal weighting method

Equal weighting describes the method of assigning an equal weight to each variable. Of course, this does not avoid weighting. The literature mentions many composite indicators that use equal weights for all indicators (Böhringer & Jochem, 2007, p. 6; Hsu et al., 2013; Nardo, Saisana, Saltelli, & Tarantola, 2005, p. 55). For example, the Human Development Index weighs the variables representing health, education, and income by one third each (United Nations Development Programme, 2014).

Figure 3: Equal weighting applications: top-down or bottom up.



Own presentation. Boxes below the phenomenon represent dimensions and their weights.

A composite indicator has multiple levels when at least one dimension is split into sub-dimensions. In these cases, equal weighting can be applied top-down or bottom-up. As shown by figure 3, in the top-down application, equal weights are assigned to the top level of the composite indicator first, e.g. 50% for the environmental and social sphere. Next, each dimension's weight is split among its sub-dimensions. If the numbers of sub-dimensions differ, the individual weights are different at the lowest level.

Equal weighting applied bottom-up counts all M' relevant dimensions at the lowest level and assigns equal weights given by $1/M'$. Consequently, the number of relevant dimensions (and their corresponding variables) determines the individual weights. In the bottom-up approach, dimensions at the top level may end up being unbalanced (OECD, 2008, p. 31).

Equal weighting does not explicitly derive weights. Instead, the same weight are assigned, which implies, that dimensions at the top or bottom level are of equal importance to the phenomenon (OECD, 2008, p. 31). Consequently, it is not an explicit process of deriving weights but the selection of dimensions and variables that determines weights.

For the case of the bottom-up application, this can be shown considering sum of all weights $W = \sum_{k=1}^{M'} w_k$ with $w \in \mathbb{R}_+$. In case of equal weights $w = w_k$, thus $W = M'w$. Consequently, each variable's weight corresponds to $w_k = W/M'$. This means, that the decision to consider M' dimensions also determines each dimension's weight.

In conclusion, equal weighting imposes requirements upon the selection process. In case of the bottom-up application, the selection of relevant dimensions at the lowest level should only include dimensions of equal importance to the phenomenon. In the case of the top-down ap-

plication, the selection of relevant dimensions at each level should only include dimensions of equal importance to the sub-phenomenon. A simple comparison of dimensions' weights may or may not raise doubts about the compliance with these requirements. This becomes particularly problematic, if the selection of dimensions and variables is data-driven. Data-driven selection hinders the consideration of requirements in addition to data-availability. Lastly, weights derived through equal weights are not trade-off weights. They may not even be importance weights, if equal importance is not considered during the selection of dimensions and variables.

If the drawbacks of equal weighting prevail, one may use other methods to derive weights. In practice, equal weighting is sometimes used as a starting point, followed by other considerations to manipulate weights afterwards (Hsu et al., 2013, p. 64). These subsequent considerations may be one of the participatory methods described next.

3.4.2 Participatory methods

Participatory methods assign weights explicitly based on the opinion of the participants. For these methods, there are two aspects to be noted in addition to the method itself. These are the participants and the instructions. I present these characteristics first, followed by two survey methods, namely the budget allocation method and the Analytical Hierarchy Process (AHP), which are repeatedly mentioned in the literature (Gasparatos et al., 2008; OECD, 2008). I conclude by citing different sources that recommend the use of the AHP.

Participants may be experts, qualifying through means of expertise or stakeholders, qualifying by being affected. Experts can be from the analysts or an external group. Parris and Kates (2003) find that popular composite indicators, assessing sustainable development, often rely on a "self-appointed group of experts" (pp. 16-17). The composite indicator literature favors the inclusion of a large group of stakeholders (Gasparatos et al., 2008; Koellner et al., 2005, p. 65; Parris & Kates, 2003; Ramos & Caeiro, 2010). The OECD handbook (2008) recommends including "a wide spectrum of knowledge and experience" (p. 96). Concerning the inclusion of the public, the authors warn that people express their concern about different dimensions, instead of relative importance with respect to a phenomenon. In conclusion, the selection of participants is a starting point of participatory methods and it is not trivial.

Participatory methods are also influenced by the instructions. If participants are asked to assess the significance of a variable, two natural counter-questions are: "With respect to what?"

and “Compared to what?” (Hardegree, 2001, p. 7). If an expert is asked to assess the importance of carbon emissions to corporate sustainability, she needs to know about the primary objective and the other dimensions of the composite indicator. It is necessary to brief the participants on the objective of the weighting exercise. Experts may judge the significance with respect to social welfare, with respect to financial performance or any other objective. If experts are instructed to estimate social costs of an impact, this implies the primary objective *evaluation of welfare effects*. If the primary objective is *prediction of financial performance*, surveys have to predict real or potential influences on a company’s cash flows. Stakeholders may also be asked to consider their own interests. This requires the participants to be a balanced representation of all interest groups.

The instructions have to reflect the primary objective in a consistent way. Different primary objectives will lead to different weights. In order to avoid redundancies, I refer to the examples that illustrate the influence of the primary objective on the identification of relevant dimensions. Just like the dependency of materiality, the degree of significance is dependent on the primary objective as well.

Lastly, the method itself determines the procedure of participatory methods. Different surveying procedures aim to ease cognitive stresses. Two common methods that can be applied to incommensurable variables are the budget allocation method and the Analytic Hierarchy Process (AHP).

The budget allocation method simply asks participants to allocate points, e.g. 100 points, to the dimensions or variables that are subject to weighting. The individual weight can be calculated by $w_k = \frac{average\ points\ awarded\ to\ k}{budget\ points}$. The method’s simplicity is its main advantage (OECD, 2008, p. 96). However, considering more than ten different dimensions, can produce inconsistent results due to cognitive stresses (Nardo et al., 2005, p. 67).

One source of inconsistency is described as the splitting effect. If dimensions are split into sub-dimensions, participants tend to assign more points (M. Weber, Eisenführ, & von Winterfeldt, 1988). Hence, this bias favors dimensions with more sub-dimensions even though the amount of sub-dimensions is not a reason of higher significance. The resulting weights are importance coefficient, because the participants do not necessarily weigh single variables against each other.

The AHP is a more sophisticated survey method to derive weights based on the participants' opinions (Munda, 2008, pp. 99ff.; Nardo et al., 2005, pp. 66ff.; OECD, 2008, pp. 96ff.). It has been applied to a variety of multi-attribute decision making problems (Munda, 2008, p. 101). T. Saaty developed it in the 1980s (R. Saaty, 1987). It combines an elaborate survey with a mathematical derivation of weights.

For the survey, the participants assess the relative importance of all variables based on pairwise comparison, i.e. one-to-one. Given M' relevant dimensions, this means that every participant must assess $\frac{1}{2}M'(M'-1)$ comparisons. Numbers from 1-9 are assigned. A pair of equally important dimensions is assigned a 1. The numbers 2-9 are assigned if the first dimension is 2-9 times more important than the second dimension. All comparisons and their reciprocals are entered into a $M' \times M'$ matrix, with the diagonal made up of 1, representing equivalent importance of the variables.

To derive weights, the matrix' eigenvector is calculated. It averages all possible combinations of weights and derives one weight for each dimension (Munda, 2008, p. 100). This approach allows the calculation of a consistency measure. The consistency measure accounts for exaggerated and even intransitive assessment. Saaty recommends to tolerate up to a 10% inconsistency (Saaty, 1987, p. 172).

The advantage of the AHP is the cognitive facilitation of comparing many dimensions. Further, it can express the consistency of the comparison and allows consequent adjustments to increase consistency. It also derives weights on the ratio measurement level (Saaty, 1987, p. 161). These weights represent trade-off ratios. Furthermore, it is able to account for multilevel phenomena (Munda, 2008, p. 101). Some authors affirm its applicability in an SR context. Krajnc and Glavič (2005) explicitly suggest its application in composite indicators to compare the sustainability of companies. Hermann et al. (2007) suggest its application to assess a company's environmental impact. Compared to the other methods for incommensurable variables, the AHP seems to be the most systematic method with an increased probability of deriving consistent weights. It is applicable in various contexts, including corporate SRs.

The disadvantage is AHP's elaborate data collection. In case of 10 variables, the AHP requires the assessment of 45 pairwise comparisons. For 20 variables, 190 comparisons need to be assessed. For 50 variables, 1225 comparisons need to be assessed. This is certainly elabo-

rate compared to the budget allocation method or the equal weighting approach. However, out of the methods proposed for incommensurable variables, it is the most precise.

To sum things up, section 3.4 discussed the meaning of weights and presented three methods to derive weights. I pointed out the disadvantages of the equal weighting method. The alternative use of participatory methods requires to choose participants and to instruct them. As different methods exist, the choice of one method should be reasoned. If participatory methods are used, the group of participants and the instructions can bias the results.

3.5 Normalization

The sections of chapter 3 discussed the theoretical framework, the set of relevant dimensions G', its corresponding variables x_k, and the set H of weights w_k. For variables, where scaling is deemed necessary, I also discussed the choice of a scaling variable y_k. In this section I assume, raw data for each variable has been collected for all N objects of analysis a_j. The next step in the construction of a composite indicator is to manipulate this data so that it can be aggregated in the last step. This is achieved by normalization.

Normalization transforms variables so that they can be compared across dimensions. They account for the different ranges and units (OECD, 2008, p.80). The normalized variables are referred to as indicators following Ott (1978). The indicators may refer to absolute values, differences, or ratios of variables.

The OECD handbook (2008) introduces normalization methods by saying: “Avoid adding up apples and oranges” (p. 27). Unfortunately, apples cannot be transformed into oranges and choosing one normalization method over another has an effect on the composite indicator’s results. Ebert and Welsch (2004, p. 281) even suggest to avoid normalization altogether, as the choice of a normalization method is arbitrary. They find that this is possible based on an ordinal interpretation of a geometric aggregation rule with all data being measurable on the ratio level[8]. As this is not the case for many ESG data, I present the characteristics of typical normalization methods and their consequences.

[8] Note that appendix 2 offers an introduction into the theory of measurement levels.

The following sub-sections present a example of normalization. Further normalization methods are introduced and discussed. A focus on the choice of reference values used by these normalization methods concludes the section.

3.5.1 Example

An example illustrates the normalization step. Afterwards frequently used normalization techniques described in the composite indicator literature are presented. Based on the criteria used to choose a normalization method I also discuss their differences and consequences.

Table 3 presents the exemplary normalization of three variables based on two different methods. The min-max-normalization transforms a variable $x_{k,j}$ (or scaled variable $\widehat{x_{k,j}}$) of company j into indicators $I_{k,j}$ based on $I_{k,j} = \frac{x_{k,j} - x_{k,min}}{x_{k,max} - x_{k,min}}$ with min/max marking the highest/smallest value among the companies. In case of CO_2e and heavy metal emissions, "the less the better" objective demands $I_{k,j} = 1 - \frac{x_{k,j} - x_{k,min}}{x_{k,max} - x_{k,min}}$ (Krajnc & Glavič, 2005, p. 554). The ranking method simply ranks the companies within each dimension and assigns the rank as a number.

The indicator values in this example are obviously very different. They cover different ranges, [0-1] or [1-5], and contain different information. After ranking, the indicators do not contain any information about the intervals between the companies with respect to the original attributes. This means that the indicators can only be interpreted on the ordinal level of measure-

Table 3: Exemplary application of two normalization methods.

Company	$\widehat{x_1}$	$\widehat{x_2}$	x_3	I_1	I_2	I_3	I_1	I_2	I_3
	$\frac{t\ CO_2e}{€\ million\ sales}$	$\frac{kg\ heavy\ metals}{€\ million\ sales}$	*% female members exec. board*	.....Min-Max.....			.....Ranking.....		
a_1	363	.297	12.5	1	0	0	1	5	5
a_2	390	.100	33.3	0	1	.44	5	1	3
a_3	385	.105	50	.19	.98	1	3	2	1
a_4	379	.218	27.5	.4	.4	.4	2	3	4
a_5	386	.267	46.3	.15	.15	.9	4	4	2

Sources: The first row values correspond to values from the German BASF 2013. CO_2e represents scope 1+2. Sources are http://www.eio.org.uk/etindex.php?page=2013_europe_300_carbon_ranking and http://bericht.basf.com/2013/en/ Other values are fictitious. Own presentation.
Remarks: All values rounded.

ment with respect the three attributes.

Additionally, the divisor in the min-max method differs for the different variables. In the case of t CO_2e per € million sales it is $x_{k,max} - x_{k,min} = 390 - 363 = 27 \frac{t\,CO_2e}{€\,million\,sales}$. The other scaling factors are $.197 \frac{kg\,heavy\,metals}{€\,million\,sales}$ and 37.5% *female members of the executive board*. Also, note that the absolute values of $\widehat{x_1}$ are quite close together, while the absolute values of $\widehat{x_2}$ are more spread out. The differences between normalization methods, which explain these differences systematically, are discussed after additional normalization methods are presented.

3.5.2 Normalization methods

Various normalization methods come into question. Table 4 presents five methods that may be applicable to SRs. They can also be combined. For example quantitative variables may be standardized first and graded afterwards while qualitative data is only graded.

The ‘ranking’ method bases on ordering all alternatives within each dimension. ‘Standardization’ and the ‘min-max’ method rescale the data based on observed data. The ‘distance to benchmark’ method rescales data based on any benchmark value. ‘Grading’ delineates categories based on quantitative or qualitative information. The ‘grading’ method is a simple form of cardinalization, which transforms qualitative information into quantitative measures (Munda, 2008, p. 69).

The notations of the normalization methods have to be adapted, if variables have a “the less the better” objective. The adaptions are self-explanatory for most normalization methods. In case of the ratio transformation distance to benchmark, the adaptation of $I_{k,j} = \frac{x_{k,j}}{x_{k,bench}}$ requires the transformation of scaled variables. Variables for emissions and resource consumption typically have a minimization objective. After scaling, the complement of the scaled variable has a maximization objective. For example, € million sales per t of CO_2e replaces t of CO_2e per € million sales.

After normalization, absolute values do not represent quantities of the original variable anymore. Gasparatos (2010) states “any concept of value is lost during the normalization” (p. 1616). Therefore, the variables are not rendered ‘comparable’ in the way discussed for the derivation of weights. Trade-off ratios cannot refer to the marginal changes of the measured attribute anymore, e.g. liter of wastewater. Instead, indicators are unitless and the actual meaning of the indicators depends on the normalization method.

Table 4: List of normalization methods.

Name	Exemplary Equation	Range	Description
Ranking	$I_{k,j} = rank(x_{k,j})$	[1–*M'*]	All companies are ordered by the variable $x_{k,j}$ of dimension k. The rank() function assigns a value equal to the position in this ranking.
Standardization	$I_{k,j} = \frac{x_{k,j} - \overline{x_k}}{\sigma_k}$	centered around 0	The difference between the variable $x_{k,j}$ and the mean $\overline{x_k}$ is divided by the standard deviation σ_k of all values in this dimension.
Min-Max or feature scaling	$I_{k,j} = \frac{x_{k,j} - x_{min}}{x_{max} - x_{min}}$	[0–1]	The difference between the company j and the worst performing company is divided by the difference between best and worst performing company.
Distance to benchmark	$I_{k,j} = \frac{x_{k,j}}{x_{k,bench}}$	centered	The variable $x_{k,j}$ is divided by a benchmark value $x_{k,bench}$.
Grading	$I_{k,j} = \begin{cases} 1 \text{ if } x_{k,j} \text{ is good} \\ 2 \text{ if } x_{k,j} \text{ is satisfactory} \\ 3 \text{ if } x_{k,j} \text{ is sufficient} \\ 4 \text{ if } x_{k,j} \text{ is insufficient} \end{cases}$	[1;2;3;4]	Grading assigns a value to a qualitative observation. The values can be letters, numbers from 1-4, percentage points from 0-100, or others.

Source: (Booysen, 2002, p. 124; OECD, 2008, p. 30).
Remarks: Notations are for variables with a "the more the better" objective. All variables $x_{k,j}$ may also be scaled variables $\widehat{x_{k,j}}$.

For example, the phenomenon environmental performance will not be an aggregate of environmental impacts after the min-max-normalization of all values. Instead, it would aggregate a company's standing in relation to its competitors based on the environmental impacts.

In the example above, $I_{1,3} = .19$ does not represent CO_2e emissions from company a_3. Instead, it measures a_3's position with respect to a_1 and a_2 based on their CO_2e emissions. Thus, the indicator's origin represents the worst performing company and not zero CO_2e emissions. One may still want to use ratios of the indicators. This is possible, if the new meaning of the indicator's origin is reflected by the new interpretation. For the example, the ratio between a_1 and a_3 after min-max-normalization with respect to CO_2e is $I_{1,1}/I_{1,3} = 1/.19 \approx 5$. Its new interpretation is: a_1 is five times farther away from the worst company than a_3. If this is the interpretation of this indicator, then the measured attribute changes and the level of measurement may not be affected by the transformation. This is important because aggregation rules imply the use of cardinal data.

The choice of a normalization method depends on properties of the data and the method itself. The composite indicator literature discusses four characteristics of the methods. These are the

reference values they use, the behavior towards outliers (OECD, 2008, p. 83ff.), the effects on the level of measurement (Ebert & Welsch, 2004), and the effect of so-called implicit weights (Booysen, 2002). As reference values need to be chosen by the SR agency, I discuss their choice in greater depth.

3.5.3 Reference values in normalization methods

A reference value anchors the range of an indicator. Ideally it is constant over time so that comparisons are possible (Welsch, 2005, p. 14). A reference can refer to benchmarks within the sample or rely on external benchmarks. Using data points from the sample hinders inter-temporal comparability because the reference will change over time. Using external benchmarks requires the identification of benchmarks, which can be difficult.

'Ranking' and 'standardization' use reference values from the sample, e.g. in the form of the ranking itself, the mean, or the standard deviation. This means that the reference values depend on the objects of analysis included in set A. The 'min-max' method identifies single reference values within the sample, i.e. the worst/best value. Whenever the normalization method relates to the set of companies, a basic axiom of measurement theory is violated, namely the independence of irrelevant alternatives (Munda, 2008, 2012; Welsch, 2005, p. 15). An irrelevant company that is added to the sample changes the references and influences the results consequently.

For the methods 'grading' and 'distance to benchmark', analysts need to identify reference values. This reference can also be based on the sample or an absolute external value.

The 'distance to benchmark' method requires a quantitative measure. This may be an industry average, the best available technology, or policy targets (Moldan, Hák, Kovanda, Havranek, & Kusková, 2004; Moldan et al., 2012). The identification of a quantitative reference can be very elaborate, as the following example shows.

The European Union identifies *best available techniques* for several industries concerning their environmental impacts. This bases on the comparison of emission values for the production of a product. Among other applications, these factors are necessary to calculate emission allowances for the carbon-trading scheme. One of 33 industries under analysis is the glass industry (Scalet et al., 2013). The *Best Available Techniques (BAT) Reference Document for the Manufacture of Glass* includes benchmark levels for eight different kinds of glass and

several different production processes. It encompasses approximately 450 pages and has taken several years to compile. This example hints at the expenditures, if SRs were to normalize data based on a quantitative external benchmark for several industries and environmental impacts.

This may be the reason why the ‘grading’ normalization is wide spread among SRs. It can deal with qualitative variables. For example, an SR agency can evaluate a stakeholder management strategy or the supply chain management. This normalization method can be very subjective, if analysts cannot find distinct characteristics to identify classes like good, satisfactory, sufficient, and insufficient. Even if distinct characteristics can be identified, this is still a delicate decision that analysts make for each variable. In fact, adjusting the qualitative reference values of the ‘grading’ method controls whether an SR is more or less demanding. Independent of the chosen variables, low requirements can lead to high scores for any company. Furthermore, individual variables may be affected differently. If requirements are low for one variable, this variable turns out to be negligible (Booysen, 2002, p. 124).

To sum things up, the normalization step includes two major decisions. First, the normalization method needs to be chosen. This choice affects the indicator values and their interpretation. Second, some methods require reference values. If these reference values can be chosen, this choice determines how demanding an SR is.

3.6 Aggregation

The last step of the construction of composite indicators is the aggregation of its constituents. Step 2, the data selection, identified a set of relevant dimensions G' and corresponding variables. Step 4 transformed these variables into indicators. The derivation of the set of weights H was described by step 3. The aim of aggregation is to combine the set of weights and the set of indicators for each company in order to derive a single measure (Munda, 2012, p. 339). An aggregation rule defines the algorithm or function leading to this one-dimensional measure. Aggregation rules can lead to results on the ordinal or cardinal levels or measurement.

The composite indicator literature discusses advantages and disadvantages of various aggregation rules (Böhringer & Jochem, 2007; Ebert & Welsch, 2004; Munda, 2012; OECD, 2008; Zhou, Ang, & Poh, 2006). The sources, which favor a single aggregation rule, base their preference on very different arguments. Zhou et al. (2006) favor the so-called weighted product

based on a measure that calculates the statistical characteristics of information content before and after aggregation. Hsu et al. (2013, p. 67) claim that the choice of an aggregation rule has to consider the understanding of non-experts if the results are supposed to be communicated to the general public. Hence, they favor the arithmetic mean. Munda (2008) considers the meaning of weights and the applicability as a measure of strong sustainability. Based on these considerations, he proposes a non-compensatory algorithm. Ebert and Welsch (2004) favor an ordinal interpretation of the geometric mean based on ratio variables to avoid normalization.

It is not surprising that different arguments lead to different conclusions about the ideal aggregation rule. Hence, the composite indicator framework does not recommend an ideal method for the aggregation, which SRs could adapt. However, the choice of one aggregation rule over the other influences the results (Esty et al., 2005, p. 37). In order to illustrate these influences, I apply two exemplary aggregation rules. I name the assumptions they make and I discuss compensability, the main difference between the various aggregation rules.

3.6.1 Exemplary aggregation rules

I chose to compare the most common aggregation rule, namely the weighted arithmetic mean, to an alternative compensatory aggregation rule, namely the weighted geometric mean (Ebert & Welsch, 2004; OECD, 2008; Zhou et al., 2006, p. 307). As SRs typically arrive at interval results, I discuss aggregation rules that derive cardinal results. Both are compensatory, which means that they assume *compensability* among the indicators. Subsequently, indicators are treated like substitutes for each other. A one-unit decrease in one indicator can be substituted by some increase in other indicators. The exemplary aggregation rules differ in the degree of compensability they assume. The assumption of compensability needs to be discussed for the concrete indicators and the dimensions they stand for (Munda & Nardo, 2005).

Given that the sum of weights $W = \sum_{k=1}^{M'} w_k = 1$, the weighted arithmetic mean is defined by:

$$SR_j = \sum_{k=1}^{M'} w_k I_{k,j} \text{ with } j = 1, \dots, N \text{ and } k = 1, \dots, M' \qquad (1)$$

Formula (1) shows that the weighted arithmetic mean is the sum of all indicators I_k attributed to a company j multiplied by the corresponding weight w_k. If the sum of weights $W = \sum_{k=1}^{M'} w_k \neq 1$, the right side of the formula needs to be divided by W.

The weighted arithmetic mean interprets weights as trade-off ratios, as it enables complete compensability (Munda & Nardo, 2005). It requires data to be normalized and measured at the interval level (Ebert & Welsch, 2004, p. 278). Additionally it assumes mutual preference independence[9] among the indicators, as it cannot account for conflicts and synergies between indicators (Munda, 2008, p. 89).

Given that the sum of weights $W = \sum_{k=1}^{M'} w_k = 1$, the weighted geometric mean is defined by:

$$SR_j = \prod_{k=1}^{M'} I_{k,j}^{w_k} \quad \text{with } j = 1, \dots, N \text{ and } k = 1, \dots, M' \tag{2}$$

Formula (2) shows that the weighted geometric mean is the product of all indicators to the power of their weighting coefficient. In case the sum of weights is $W = \sum_{k=1}^{M'} w_k \neq 1$, the right side of the formula is subject to the W-th root.

The weighted geometric mean reduces compensability among its indicators compared to the weighted arithmetic mean. It requires data to be normalized and measured at the ratio level (Zhou et al., 2006). Note that this is depending on the interpretation of the indicators.

Taking the example from section 3.5.1 again, table 5 presents the results of applying (1) and (2) to the indicator values after the min-max normalization. To facilitate the observation of differences, let us assume equal weights $w_1 = w_2 = w_3 = \frac{1}{3}$.

Table 5 indicates differences between the aggregation rules. Companies a_1 and a_2 are punished by the weighted geometric mean for being the worst performing companies in at least one dimension. Note that the application of the weighted geometric mean interprets the indicator values as ratio-variables. This means that the range between the worst and best performing company has a true zero. Furthermore, a_4 and a_5 receive the same score with the weighted arithmetic mean, while the weighted geometric mean penalizes the imbalance of a_5.

[9] Mutual preference independence is defined by (adapted from Munda, 2008, p.89): A subset of indicators G'_{sub} is preferentially independent of its complement G'_{comp} only if any conditional preference among elements of G'_{sub}, holding all elements of G'_{comp} fixed, remain the same, regardless of the levels at which G'_{comp} are held. The indicators are mutually preferentially independent if every subset G'_{sub} of these indicators is preferentially independent of its complementary set.

Table 5: Exemplary application of aggregation rules.

Company	I_1	I_2	I_3	CI	CI
	……Min-Max……			weighted arithmetic mean	weighted geometric mean
a_1	1	0	0	.33	0
a_2	0	1	.44	.48	0
a_3	.19	.98	1	.72	.57
a_4	.4	.4	.4	.4	.4
a_5	.15	.15	.9	.4	.27

Own presentation.

To sum things up, the order of the firms changes depending on the aggregation rule. According to the weighted arithmetic mean $a_3 > a_2 > a_4 = a_5 > a_1$ and according to the geometric mean $a_3 > a_4 > a_5 > a_2 = a_1$. This is because the assumed compensability differs. The next section discusses the different degrees of compensability and the consequences for the meaning of weights.

3.6.2 Compensability and the meaning of weights

In order to see how the arithmetic and geometric aggregation rules affect the assumed compensability, we can derive the rate of substitution for (1) and (2). The formal derivation is shown in appendix 3. The rate of substitution $S_{1,2}$ is defined as the marginal utility of I_1 divided by the marginal utility of I_2. It describes how much I_2 needs to change in order to compensate for a given change in I_1. For the weighted arithmetic mean, the rate of substitution $S_{k,r}$ between any two indicators I_1 and I_2 is given by:

$$S_{1,2} = \frac{w_1}{w_2} \tag{3}$$

For the weighted arithmetic mean, $S_{1,2}$ is equal to the ratio of the weights. This is the reason why the weights have the meaning of trade-off ratios (Munda & Nardo, 2005). Further, the rate of substitution is constant and assumes an infinite elasticity when one indicator is approaching zero. Therefore, deviations from the average by some indicators can be compensated by other indicators without losses.

In the case of the weighted geometric mean, the rate of substitution is given by:

$$S_{1,2} = \frac{w_1}{w_2}\frac{I_2}{I_1} \tag{4}$$

The geometric aggregation leads to a rate of substitution proportional to the inverse ratio of the indicators. Subsequently, the elasticity is constantly 1 because the ratio of relative changes does not change. If the $I_1 = I_2$, the weights are interpreted as trade-off ratios as well. However, the ratio changes. In the extreme case, if an indicator approaches zero, the rate of substitution approaches infinity. Therefore, deviations from the average by some indicators lead to lower scores compared to a set of indicator values that is collectively close to the mean. This is also the reason why, if an indicator approaches zero, the value of SR_j approaches zero as well, if it is defined at all[10]. This illustrates the weighted geometric mean's assumption about the origins of indicators. It requires data measured on the ratio level. After an interval transformation, this requires a new interpretation of the origin.

Both functions show, that the weights influence the rate of substitution (Munda & Nardo, 2005). Reversely, this affects the derivation of weights. First, methods need to be used to derive trade-offs. Second, these trade-offs need to apply to the indicators. In section 3.4 about the derivation of weights, I referred to variables. However, the real trade-off happens between indicators. The normalization is interposed between the selection of variables and their aggregation. In consequence, the weights apply to indicators and this meaning of weights needs to be reflected during the derivation of weights.

The following example illustrates the implications for the derivation of weights. Consider a weighted arithmetic mean after min-max normalization for CO_2e and wastewater. If participants assign $w_{CO_2e} = .15$ and $w_{ww} = .3$, the rate of substitution would be $S_{CO_2e,ww} = 2$ according to (3). If we take into account the normalization method, this means that being 1% closer to the best performing company in with respect to CO_2e compensates for being .5% farther away from the best performing company with respect to wastewater. The trade-off ratio does not apply to absolute amounts of CO_2e and wastewater anymore. In practice, the required abstractive power of participants leads to difficulties regarding the correct derivation

[10] Different conventions exist. Occasionally the geometric mean is not defined for zero.

of weights (Munda, 2008, p. 91). A “theoretical inconsistency” exists, if weights are derived as importance coefficients and are applied in compensatory aggregation rules (Munda & Nardo, 2005, p. 7).

To conclude the comparison of the weighted arithmetic mean and the weighted geometric mean, the main difference concerns the compensability. While an additive aggregation assumes full and constant compensability at all levels, the multiplicative aggregation limits the compensation by reflecting relative changes. It penalizes companies with differences between dimensions, e.g. a_5 in the introductory example, compared to companies like a_4 that perform equally well in all dimensions. Thus, companies with weak performance in some dimensions should be interested in the use of the weighted arithmetic mean. At the same time, marginal increases at a low level are rewarded more than marginal increases at a high level. Consequently, the use of the weighted geometric mean implies that companies are supposed to take equal care of all sustainability dimensions and leadership in some dimensions is not sufficient.

One could argue that if the *prediction of financial performance* is the primary objective, then the arithmetic mean should be used, as financial risks are completely compensatory. If the *evaluation of welfare effects* is the primary objective, then one may tend to use the geometric mean, particularly if low performance in some dimensions outweighs high performance in other dimensions.

There is a strong argument to assume compensability of environmental indicators in the case of companies. Their economic activities are a manifestation of the division of labor. Division of labor is necessary to use comparative advantages and economies of scale. At the same time, the division of labor leads to unbalanced environmental impacts because the specialization of a company determines the its impacts. For example, along one value chain, forestry operations cause very different environmental impacts than the pulp and paper industry and printing. This is an inevitable consequence of the division of labor. Non-compensatory approach do not account for that. Thus, it seems justified to use compensatory approaches for the assessment of corporate *environmental* sustainability.

The discussion is less clear concerning social aspects of sustainability. Can a company with great support mechanisms for minorities at their European headquarters be lax about child labor in their supply chain? A normatively oriented consideration may deny. A purely eco-

nomic approach may estimate the reputation risks of child labor. The same difficulties apply when social and environmental aspects are considered together.

There are three options to account for concerns about compensability. One is the implementation of exclusion criteria, as mentioned in chapter 2. These work as lexicographic filters outside of the composite indicator framework (Munda, 2008, p. 4) and exclude any compensability for those criteria. The second option is increasing the relative weight of the dimensions that are considered more important. This makes it more costly to compensate for one dimension. Lastly, one may choose a geometric aggregation, e.g. the weighted geometric mean. This effectively excludes compensability for dimensions where a company scores zero.

To sum things up, for the methodical choice of the aggregation rule, there is no correct way. Full compensability is not self-evident. However, restricting compensability by using the weighted geometric mean brings along other implications. For example, the origin is interpreted as a true zero, which assumes a ratio level of measurement. Both aggregation rules are applicable and the choice of one of them affects the results of the SR.

3.7 Remarks

It is tempting to split the reviewed literature on composite indicators into two general groups: practitioners (Esty et al., 2005; Galli et al., 2008; Hsu et al., 2013; Krajnc & Glavic, 2005; Nardo et al., 2005; OECD, 2008) and theorists (Böhringer & Jochem, 2007; Booysen, 2002; Ebert & Welsch, 2004; Gasparatos et al., 2009; Munda, 2005, 2008, 2012; Parris & Kates, 2003). While theorists insist on a theoretically sound methodology, practitioners have (implicitly) accepted that many compromises need to be made when multi-dimensional phenomena are aggregated into one measure. The two groups differ concerning the extent to which they accept methodical decisions to be mere value judgments where there is no right or wrong. Of course, this is a rough differentiation. Still, it can be noted that the literature does not agree on acceptable compromises in the construction of composite indicators.

Yet, *all* authors agree that the methods used and decisions made during the construction of composite indicators need to be transparent to enable the interpretation of the composite indicator. Based on this, the next chapter derives transparency criteria by summarizing the crucial parts of the construction.

4 Definition of Transparency Criteria

Chapter 3 explained the construction of an SR based on the composite indicator framework. I presented steps that influence the results directly and described alternative methods along with difficulties and implications of using one method or another. This chapter is based on these insights to set up transparency criteria. It tackles the first research question:

(RQ1) *What information is necessary to interpret a one-dimensional measure that represents a multi-dimensional phenomenon like corporate sustainability?*

The transparency criteria constitute a checklist of information that needs to be published about the construction of an SR in order to adequately interpret it. I claim that if one of these criteria is not fulfilled, one cannot interpret the results of an SR. An application of these criteria to various SRs is demonstrated in chapter 5.

Note that the transparency criteria concern the construction of an SR. A SR agency should also be transparent about other areas, e.g. its institutional setup, financial dependencies, and its quality management processes. These aspects are not covered by the transparency criteria concerning the construction of SRs.

4.1 Reasoning of Transparency Criteria

The composite indicator framework structures the construction of a one-dimensional measure to measure a fuzzy phenomenon. From this structure, five steps directly influence the resulting scores of the SR. These are the theoretical framework, the data selection, the derivation of weights, the normalization, and the aggregation. This section extracts the lessons learnt from this research for the construction of an SR. It describes elements that are necessary to construct and interpret an SR. Each paragraph recapitulates the element and is formulated to derive a transparency criterion. The argument is always the same: If there is more than one alternative to construct the SR and if this decision influences the resulting SR scores, then it is necessary to describe this element in order to enable the interpretation of the SR. The following paragraphs derive the necessary elements, before I list the transparency criteria in section 4.2.

The theoretical framework builds the basis for the construction of an SR. It does not comprise any methods. Instead, it names three elements that are the basis for the construction of the SR.

First, the primary objective describes the purpose of the SR. Second, the objects of analysis are what the measurement is applied to. They are described by the set of companies and the scope of the individual company. Third, the phenomenon is what is being measured. This is described by a definition of corporate sustainability. In the likely case of a multilevel definition, sub-phenomena need to be defined, too. Further, I proposed to spell out the definition using a global set of definitions G with M dimensions that are mutually exclusive and collectively exhaustive. The global set of dimensions can be defined independent of the primary objective. Any SR is built on the assumption that measuring corporate sustainability contributes to achieving its primary objective. All three elements influence the SR scores because of they are the basis of other steps of the construction process. Hence, the definition of an SR's primary objective, the set of companies, the scope of the individual company, the definition of corporate sustainability need to be described to enable the interpretation of an SR. In addition, I propose to disclose the global set of dimensions so that stakeholders can agree on a phenomenon's dimensions independent of their objectives.

The data selection is a two-fold process. First, the selection of relevant dimensions is based on the definition of the phenomenon and guided by the selection criterion. The selection criterion spells out the primary objective and should be as precise as possible to select those dimensions that are actually considered by the SR. Even if the selection criterion is precise, the selection of dimensions relies on case-by-case decisions. Second, the selection of variables concludes with M' variables x_k that operationalize the relevant dimensions m_k of set G'. The selection of variables is guided by quality characteristics of this data. Scaling variables may complement the variables, e.g. to account for different company sizes. The selection of variables and scaling variables are ultimately case-by-case decisions as well. All parts of the data selection influence the results because they identify what is actually measured. As analysts cannot choose one method, disclosure of the individual dimensions and variables is necessary. Hence, the selection criterion, the set of relevant dimensions, the selected variables, and the scaling variables need to be described to enable the interpretation of an SR.

The derivation of the set of weights H constitutes the second input argument to the SR's aggregation rule. It can use different methods to derive set H with M' weighting coefficients w_k. The method describes a procedure that allows quantifying the significance of variables to the phenomenon (importance) or their significance with respect to each other (trade-off ratios). Different methods lead to different weights. Thus, the interpretation of weights depends on the

method that was used to derive the values. However, even if the method is clearly spelled out, it does not eliminate biases and room for interpretation. Thus, the derivation of weights can only be understood by a transparent disclosure of the derived weights themselves. In conclusion, the method to derive weights and the set of weights need to be described to enable the interpretation of an SR.

Participatory methods to derive weights are only conclusively described with information about the participants and the instructions they received. Participants may be experts or stakeholders. The instructions need to be consistent with the primary objective. Both influence the resulting weights and in turn the SR scores. Hence, the participants and instructions need to be described to enable the interpretation of an SR if a participatory method is used to derive weights. In other words, these are contingent requirements, namely contingent on the case that a participatory method is used to derive weights.

Normalization prepares the variables for aggregation. The normalization method describes a transformation of variables to account for differences between dimensions. I discussed alternative normalization methods and illustrated that they yield different indicator values with different interpretations. Some of these methods are based on reference values. They anchor the range of the resulting indicator or serve as an evaluation criterion. Normalization methods that rely on reference values are only conclusively described with information about these reference values. The calculated indicators enter the aggregation rule and have a direct influence on the SR scores. To sum things up, the method of normalization needs to be described to enable the interpretation of an SR. Contingent on the case that reference values are used during normalization, these reference values need to be described to enable the interpretation of an SR.

Lastly, the aggregation compiles all indicators I_k of one company j and all weights w_k based on an aggregation rule. The aggregation rule defines the algorithm or function leading to a single measure. I discussed two aggregation rules and illustrated that they can change the SR scores and the ranking of the companies. Thus, the choice of one aggregation rule influences the results. There is no agreed method to choose the correct aggregation rule in the composite indicator literature. In conclusion, the aggregation rule needs to be described to enable the interpretation of an SR.

4.2 Transparency Criteria

In the section above, I derived elements of each step in the construction of an SR that are necessary pieces of information for the interpretation of an SR. This section lists the 14 transparency criteria. A transparency criterion is a test that checks whether an element has been published or not. The set of transparency criteria enables the evaluation of an SR's transparency. Based on all transparency criteria, one can state whether sufficient information is available to interpret an SR. If the information is not sufficient, the transparency criteria point out areas where more information is necessary.

The transparency criteria have all been derived from the composite indicator framework with one exception. As mentioned in chapter 2, SRs use exclusion criteria. These work as lexicographic filters outside of the composite indicator framework (Munda, 2008, p. 4). They restrict the sample of companies by excluding companies that do not pass these filters. This may also affect the results of relative normalization methods because they depend on the sample. Consequently, they are included in the transparency matrix as one of three elements describing the objects of analysis.

Further, note that the transparency criteria do not demand to publish any collected data. The construction and interpretation of an SR is independent of the values put into this algorithm. This data is what customers pay for and it would be pointless to publish it free of charge. Consequently, the transparency criteria include neither the variable values for all companies, nor the indicator values, nor the SR's results.

There are two levels of transparency criteria. The first level comprises elements of the construction that must be published to enable interpretability of the SR. The second level comprises elements that are only required contingent on the case that a certain method is chosen. This second level complements the first level. In general, the first level asks for the choice of method and the second level asks for details of a specific method. Concerning the theoretical framework and the data selection, there are no alternative methods. One needs to define the elements of the theoretical framework explicitly.

Furthermore, there are three transparency levels for each criterion. The first one is the reference to the demanded element. It is not sufficient to refer to an element without describing it. A reference would be "we apply weights" without describing which weights are applied. Hence, a reference only acknowledges the existence of an element, but it does not entail the

Table 6: Transparency criteria for the construction of SRs.

#	Element	Reference	Description	Justification
Theoretical Framework				
TC1	Definition of the primary objective	Yes/No	Yes/No	Yes/No
	Objects of analysis			
TC2	Exclusion criterion	Yes/No	Yes/No	Yes/No
TC3	Set of companies	Yes/No	Yes/No	Yes/No
TC4	Scope of individual company	Yes/No	Yes/No	Yes/No
	Phenomenon			
TC5	Definition of corporate sustainability	Yes/No	Yes/No	Yes/No
TC6	Global set of dimensions[1]	Yes/No	Yes/No	Yes/No
Data selection				
TC7	Selection criterion	Yes/No	Yes/No	Yes/No
TC8	Set of relevant dimensions	Yes/No	Yes/No	Yes/No
TC9	Selected variables	Yes/No	Yes/No	Yes/No
TC10	Scaling variable	Yes/No	Yes/No	Yes/No
Derivation of Weights				
TC11	Method to derive weights	Yes/No	Yes/No	Yes/No
TC11a	Participants[2]	Yes/No	Yes/No	Yes/No
TC11b	Instructions[2]	Yes/No	Yes/No	Yes/No
TC12	Set of weights	Yes/No	Yes/No	Yes/No
Normalization				
TC13	Method for normalization	Yes/No	Yes/No	Yes/No
TC13a	Reference values[3]	Yes/No	Yes/No	Yes/No
Aggregation				
TC14	Aggregation rule	Yes/No	Yes/No	Yes/No

Own presentation. This table lists the transparency criteria. Its interpretation is "TC4 can be fulfilled on three levels. If the scope of individual company is referenced, it fulfills the reference level for TC4. If the scope of individual company is described, it fulfills the description level for TC4. If the scope of individual company is justified, it fulfills the justification level for TC4."
The first column numbers the transparency criteria for referrals in the following sections.
Remarks: [1] Optional suggestion. [2] Only applicable for participatory methods. [3] If method for normalization uses reference values, e.g. best available technology.

content necessary to interpret an SR. The second level is the description of the element. When an agency describes an element, the reader knows which method is used or which decision was taken. If all elements of the transparency criteria are described, the collective information is sufficient to interpret the SR.

The third level is the justification of the respective element. This level is ideal because many decisions are debatable and an agency may forestall doubts about its SR by justifying it on the

basis of its primary objective. At the same time, the need to justify an element differs. For example, the set of companies is simply a question of demand. The SR assesses those companies customers ask for. Thus, there is little need for justification. On the other extreme, much more justification is needed regarding the derivation of weights. However, the interpretation of an SR is possible based on the set of weights.

In conclusion, the justification is *not necessary* to enable the interpretation of an SR. At the same time, a reference to an element without describing it is *not sufficient* to enable the interpretation of an SR. This seems trivial, but it needs to be noted that referring to a normalization methods does not disclose information about the normalization method, referring to weights does not disclose information about the weights, referring to aggregation does not disclose information about the aggregation rule, and so on. The *descriptions* of the elements of the transparency criteria are necessary pieces of information for the interpretation of an SR. If all elements of the transparency criteria are *described*, the collective information enables the interpretation of an SR.

The transparency criteria in table 6 are mostly self-explanatory. Few notes shall clarify them. The following paragraphs mostly refer to the description of these elements.

The theoretical framework needs to be described by a primary objective (TC1), the objects of analysis and the phenomenon. The objects of analysis are defined by the exclusion criteria that are used to pre-select the universe of relevant companies (TC2), the companies that are actually considered by the SR (TC3), and the scope of each individual company that is considered (TC4). The set of companies does not have to be a list of the individual companies. It is sufficient to describe characteristics such as an index, the company size, the industry, and the geographical origin. For the phenomenon, an agency should publish how it defines corporate sustainability in general, disregarding its own primary objective (TC5). It may do so by laying out the global set of definitions it considers (TC6).

Four parts describe the data selection. The guiding selection criterion (TC7) should be described. The set of dimensions (TC8) that are relevant regarding the primary objective of the SR and which the SR consequently considers, should be listed. Further the selected variables (TC9) and scaling variables (TC10) should be assigned. These four parts need to be described in a way so that it is clear how they belong together. In other words, the reader needs to be enabled understand which dimension is measured by which variable and whether this variable

is scaled. A justification of these elements refers to the whole sample and explains the basis on which the individual variables were chosen.

Two elements describe the derivation of weights. First, the method to derive weights (TC11) should be explained by the procedure that leads to the actual values. This transparency criterion has two contingent criteria. Namely, a description of the group of participants (TC11a) and the instructions (TC11b) need to be described contingent on the case that a participatory method is used to derive weights. Second, the set of weights (TC12) illustrate the importance of each dimension for the SR.

A description of the method of normalization (TC13) is sufficient to interpret the indicators after normalization. Again, this criterion has a contingent criterion. Contingent on the case that the method for normalization uses reference values that cannot be identified from the sample, these reference values (TC13a) need to be described.

Lastly, the aggregation is described by the aggregation rule (TC14). This is achieved by a function or the term of the aggregation rule.

4.3 Justification of Transparency Criteria

One may question any individual criterion. In section 4.1, I showed that there is at least more than one possible way to go about these elements and that the necessary decision influences the result. The justification of the transparency criteria shall anticipate two kinds of general criticisms. One can question the transparency criteria's addressees and the selection of criteria. Colloquially speaking this corresponds to "why is transparency needed at all?" and "why these criteria and not others?".

Possible groups of addressees are the customers, the companies, and the public. In general, axiom 2 states that an SR agency has the obligation to be transparent to a degree sufficient to interpret it. I made this axiom comprehensible by listing stakeholders that are interested in interpreting the SR. I did not justify the individual stakeholder's claims. The composite indicator literature demands full disclosure to the public. However, composite indicators typically address policy makers and the public. Hence, this argument is not applicable to SRs. In the case of SRs, a formal, consequential derivation of the legitimate addressee of transparency would have to consider the societal role of SRs and their license to operate. A less formal ar-

gument for public disclosure may note that SRs depend on the transparency by companies and should set a good example themselves (Arise, 2012a). In addition, Sadowski, Whitaker, and Ayars argue that transparency creates trust in SRs, which in turn leads to higher adoption by customers (2011). Hence, there may be a business case for transparency to the public.

If one does not agree that the public should be able to interpret an SR, two other addressees can be considered. Naturally, the group of customers has a well justified claim to full transparency as they hesitate to finance the SR, if they paid for a number or grade only. They need to be enabled to interpret and apply the SR. Further, the group of rated companies may also have a well justified claim to know how it is evaluated. The currently developed *Sustainability Ratings Standard* by the Global Initiative for Sustainability Ratings (2013) includes a similarly restricted definition of transparency as the first of twelve principles: "A rating should be transparent to those whose decisions are affected by the application of such rating" (p. 9).

Ultimately, I conclude that the interpretation of the results requires transparency. A SR agency should be transparent to all those with well justified claims to interpret the results. The thesis does not identify the group that has such claims to interpret the results beyond doubt. However, there are strong intuitive arguments in favor of full disclosure to the public. The burden of proof is clearly in the field of those trying to argue against transparency requirements.

The second criticism questions the concrete selection of the transparency criteria. They may seem arbitrary, given the possibility that an SR agency may not even have defined its primary objective internally. Indeed, the justification depends on two assumptions because of this thesis' approach.

First, the composite indicator framework needs to be applicable to SRs. In chapter 2 I argued for this assumption. I showed that the composite indicator framework and SRs share key characteristics, constituents, and have an equivalent problem setting. Thus, the framework and the methodology are applicable to SRs and I concluded that structure and findings of this research are valid for SRs as well.

Second, one needs to assume that the five steps of this thesis are the relevant and correct steps for the interpretation of a one-dimensional measure. The literature discusses other steps in addition to the five steps. For example, one could include the publication of an uncertainty and sensitivity analysis in the transparency criteria. The results would describe the reliability and robustness of the results. I excluded this and other steps, because they do not influence the

results directly. One may argue for the inclusion of additional transparency criteria. However, they cannot be supported by the attempt to interpret an SR.

An alternative justification is applied in section 4.1. For each criterion, one may ask how an SR can be constructed without its explicit formulation. If the set of transparency criteria is justified, each criterion needs to be necessary for the construction. For example in the absence of an explicit primary objective, there is no decision criterion to choose relevant dimensions. It is also not possible to derive weights without defining what these weights relate to. In reverse, this means that the addressee cannot replicate and interpret the SR.

So sum things up, transparency is needed for the interpretation of an SR. The addressees with well justified claims may be contested. The concrete selection of the proposed transparency criteria is based on the applicability of the composite indicators framework and the individual element's influence on the SR scores.

4.4 Remarks

The transparency criteria represent the end of the theoretical parts of this thesis. The objective was to reflect the construction of SRs based on an established academic background. I identified the composite indicator framework as an applicable approach. The discussion of the construction of composite indicators enables many conclusions about SRs. I presented five steps that are crucial for the construction of SRs and I referred to specific circumstances of SRs where applicable. In this chapter, I used these findings to set up transparency criteria. The elements that are required by these criteria are the elements that are used for the construction of an SR. Users of SRs can only interpret the results, if this construction is replicable and comprehensible.

Two facts are important about the transparency criteria. First, compliance with these criteria enables interpretation. I did not identify the group of addressees that is entitled to interpret SRs. I referred to arguments in favor of public disclosure, but I refrained from a single conclusion. Second, the transparency criteria cover the construction of one-dimensional measure that determines any SR according to my definition. In addition, there may be further arguments for an SR agency to be transparent about other areas, too. This is not part of this thesis.

5 Review of Sustainability Ratings

This chapter demonstrates the application of the transparency criteria. It tackles the third research question:

(RQ3) *Do SR agencies publish sufficient information for the interpretation of the SR results?*

The review is carried out for five SR agencies. They published so-called transparency matrices for being accredited by the Responsible Investment Research Standard Arista 3.0 in 2012. This highly regarded standard suggests that these five agencies are particularly transparent about the construction of their SR. The review still needs to assume addressee and medium. Namely, SR agencies need to be transparent to the public and they disclose the corresponding information through the Arista 3.0 transparency matrices. This first assumption has been discussed sufficiently. A presentation of Arista 3.0 will make the latter claim plausible.

This chapter follows the classic structure of empirical analyses. After the hypothesis, I present findings from the literature. I introduce the sample by describing the specificities of Arista 3.0. The results are presented in form of two tables and corresponding citations. A discussion of the results is concluded by final remarks.

5.1 Hypothesis

My hypothesis for this review is:

(H1) *The information disclosed by SR agencies publicly does not allow the interpretation of the SRs' results.*

In other words, the public is not informed about what is actually measured by the reviewed SRs. I expect the hypothesis to be true for several reasons. First, the transparency criteria demonstrate that much specific information is needed to interpret the results of an SR. Second and more specifically, I expect little specificity about the implementation of the primary objective because there is an incentive to serve multiple audiences. This due to the combination of high fixed costs for research and low unit costs for the access to the SR. A SR agency can increase profits by pretending to serve the objectives of more than one audience. These audiences may have different primary objectives. As demonstrated in chapter 3, an SR can only serve one primary objective. By publishing this one primary objective transparently, an SR

agency would reduce potential customers and profits. Third, agencies may try to avoid a debate about methodological decisions. For example, the derivation of weights is a delicate step, which is likely to cause controversies. The same is true for the normalization of variables by grading qualitative observations. Fourth, SR agencies may see their methodology as intellectual property. The fear of copycats usually prevents innovators from publishing their approach. By not publishing details, this risk is limited.

5.2 Findings in the Literature

No study looks explicitly at SRs' transparency regarding the construction of the one-dimensional measure. However, studies, which analyze SRs for other reasons, find that (1) SRs differ from each other and that (2) the current transparency is not sufficient to comprehend the different results.

Schäfer et al. (2006) draw a differentiation between economically motivated and normatively motivated approaches. This is reflected by the different "analytical schemes and criteria" due to "providers' own individual motives and perhaps very different notions of sustainability" (p. 16). Döpfner and Schneider (2012) confirm these findings and stress the need of SRs to disclose their methodology.

Unfortunately, the motives and objectives are not spelled out by SR agencies as shown by Chelli and Gendron (2012). They analyzed the language used by SR agencies on their websites and reports and find that SR agencies promote an 'ideology of numbers' through different discursive modes implicitly. The analysis concludes that SR scores are a stylized version of a company's social responsibility without clearly stating their intentions.

The need for transparency has also been shown by Delmas and Blass (2010). They assess fifteen chemical companies with different sets of variables. The results vary widely because the companies with "lower environmental performance and compliance tend to provide better quality of environmental reporting and to adopt more pollution prevention activities" (p. 256). Similar reasons have lead other authors to criticize the lack of transparency without further specification (Delmas et al., 2013; Dillenburg et al., 2003; Windolph, 2011).

An extensive review has been carried out by the think tank SustainAbility between 2010 and 2012. Their project "Rate the Raters" polled financial analysts, SR agencies, and companies to

find out about the status quo, the standing, challenges, and best practices of SRs. They find that "the sauce [of SRs] remains secret" (Sadowski et al., 2010, p. 5). In a follow-up publication Sadowski, Whitaker, and Ayars (2011) report more differentiated results based on a review of 21 SRs. They note that "nearly all ratings cite some variation of 'improving corporate performance and/or transparency on sustainability issues' as their objective" (p. 7). In their opinion, these objectives are too general and do not add enough value to the audience. They recommend that SRs should narrow their audience to enable clearer objectives. Further, they find strong disclosure practice of the methodology by a majority of the SRs to at least one target audience, but not necessarily to the public. They define a strong disclosure as follows:

> A rater fully discloses its methodology to the public, including its selection process, information sources, criteria, areas of evaluation, scoring schemes, assumptions and rules. This information allows the user to fully understand and replicate how the rating is constructed. (Sadowski, Whitaker, & Ayars, 2011, p. 11)

Note that this definition roughly corresponds to my transparency criteria in a less formal way. Areas of evaluation are dimensions, criteria are variables, scoring schemes may refer to weights and the grading normalization. Assumptions and rules are omnipresent, but may correspond to the aggregation rule.

Finally, the publication questions the complexity of some SRs. Too numerous variables hinder the understanding. The analyzed SRs use between 20 and 700 different indicators. Lydenberg et al. (2010, cited in Sadowski, Whitaker, & Ayars, 2011, p. 25) present a method to select material indicators. They arrive at less than 30 indicators per industry. This indicates that the disclosure of variables is may be connected to the quality of the SR. An agency that has not identified the material dimensions is unlikely to disclose its collection of dimensions.

Several authors refer to the quality standard Arista 3.0 and its precursors as an example of increased transparency (Döpfner & Schneider, 2012; Eurosif, 2014; Global Initiative for Sustainability Ratings, 2013; Novethic, 2013; Sadowski, Whitaker, Lee, & Ayars, 2011; Schäfer et al., 2006; Windolph, 2011). In their response to the Rate the Raters' survey, three SR agencies (Oekom, Vigeo, and Eiris) refer to their Arista 3.0 transparency matrix for a description of their methodology. Arista 3.0 and the sample for this review are introduced next.

5.3 Introduction of the Responsible Investment Research Standard Arista 3.0

An industry-initiative brought forward the most comprehensive quality standard for SRs. So-called transparency matrices document its accreditation process. These are published online, which makes them an accessible resource for this empirical review.

The Responsible Investment Research Standard Arista 3.0 is owned and managed by the Association for Responsible Investment Services (Arise). Arise has thirteen members that have signed the Arista 3.0 comprising eleven quality principles and nine integrity principles. Eleven of them were externally audited. It has a history of twelve years. The European Commission supported the development of its precursors in 2002. It was then known as the Voluntary Quality Standard for Corporate Sustainability and Responsibility Research (CSRR-QS). The Arista 3.0 is the third version of this standard, implementing a broader scope and a new branding since 2012.

One of the objectives of Arista 3.0 is to "stimulate transparency" about the "specificities of [SR agencies'] assessment processes" (Arise, 2012a, p. 3). The SR agencies commit to "be transparent about the methodology (the range of criteria used, the involvement of stakeholders, and the coverage) and to avoid or make transparent 'black box' approaches where it is unclear how the assessment/rating result is achieved" (p. 5). The accountability principle of Arista 3.0 commits agencies to be transparent "to the same extent as they ask companies to be transparent" (p. 6). This extent is explained in an extra section on public disclosure:

> The [SR agency] shall disclose publicly all the information needed for users and stakeholders to understand its research and evaluation methodologies in a general way, including the criteria, ratings, grades, symbols, etc. The [SR agency] should make its research methodology including any grading and ratings systems available to clients, companies, stakeholders, and the public. Should it in specific, exceptional circumstances not do so, it must make public its reasons for not doing so. (Arise, 2012a, p. 10)

The statement confirms this thesis' approach. Arista 3.0 asks SR agencies to enable the interpretation of their research and evaluation. Further, it defines the addressees of the transparency and establishes a "disclose or explain" policy should the SR agency not disclose the information to the public. In conclusion, the standard has a similar goal like the transparency

criteria of this thesis. Hence, the documentation of its auditing process should cover the same elements.

While SR agencies can be signatories of Arista 3.0 without any accreditation, there is an accreditation process for audited agencies. The objective of an audit is "to obtain and to maintain the Arista 3.0 certificate, proving compliance with the standard and showing a true and fair description of its methodology in its Transparency Matrix" (Arise, 2012a, p. 16). This matrix is "a disclosure tool for methodologies and research processes and group features" (p. 2). In practice, the transparency matrix is an Excel template to be filled out by the applicant SR agencies. An external auditor verifies "each and every line" (Arise, 2012b, p. 2) of the transparency matrix and the Arise certification council accredits a certificate of conformity with the responsible investment research standard Arista 3.0 for three years. The content of the transparency matrix encompasses the "specificities, specialties and diversities of one methodology of one [SR agency]" while it "is not rating the methodologies and is not considering that specific processes are better than others" (p. 2). The sections of the transparency matrix template are shown in table 7.

Table 7: Overview of the Arista 3.0 transparency matrix template.

#	**Name**	**Questions**	Content
A	General information	43	Facts about the SR agency.
B	Integrity and professionalism	26	Possible conflicts of interest and handling thereof.
C	Quality Management Systems	30	Responsibilities, setup, and frequencies of measures to ensure high quality.
D	Research Methodologies *	42	Overview of research activities and inspiration.
E	Research Framework *	39	Data selection.
F	Research Process *	69	Data collection including data sources, objects of analysis, and frequencies.
G	Assessment Process *	17	Assessment of data, incl. evaluation and weighting.
H	Pro-active stakeholder involvement *	20	Engagement with agency's stakeholders.
I	Products and services	29	Description of products and services from a market perspective. SR may be one of these.
J	Country rating	49	Separate methodology that is applied to countries.
K	Screening of transnational institutions	49	Separate methodology that is applied to transnational institutions, e.g. development banks.
L	Negative Screening *	44	Products based on exclusion criteria only.
M	Engagement and proxy voting services	38	Services where agency acts on behalf of its customers.
Z	Disclosure and communication	36	List of documents and addressees thereof.

Source: (Arise, 2012b). Own presentation. These are the 14 sheets of the transparency matrix. The letters are adapted. The numbers indicate the amount of closed and open questions for each sheet. There are 531 questions overall. Descriptions are mine. * marks sections relevant to the construction of SRs.

Judging by the list of contents, the agencies should describe elements of this thesis' transparency criteria in their transparency matrix. The scope is broader with the inclusion of institutional aspects, the quality management system, and related products. This review focuses on sheets D, E, F, G, H, L, and Z of the transparency matrices. The sheet's letter and the question's number identify specific questions. For example, F63 refers to the 63rd question on sheet F. The transparency matrix has open and closed questions, i.e. allowing only yes/no answers. Many of the open questions are not specific, so that an agency can influence its level of transparency on its own.

Eleven SR agencies have been accredited based on their transparency matrix as of December 2014. Most of their certificates end in February of 2015. They are listed in table 8.

Out of the eleven SR agencies, seven use the same methodology. They all belong to the research network Eiris, which is headquartered in London. The seven agencies carry out SRs based on the Eiris methodology in their geographical zones and feed the results into a global database. As their transparency matrices are identical for the methodological parts, I review the transparency matrix of the British agency only. Therefore, I am left with five SR agencies and five different methodologies.

Table 8: Sustainability rating agencies certified according to Arista 3.0.

#	Name	Abbreviation	Country	Certified Methodology	# Analysts	# companies
1	Corporate Analysis Enhanced Responsibility	Caer	Australia	Eiris	5	3000
2	Fundación Ecología y Desarrollo	Ecodes	Spain	Eiris	4	144
3	Ecovalores	Ecovalores	Mexico	Eiris	2	80
4	Ethical Investment Research Services *	Eiris	UK	Eiris	40	3000
5	EthiFinance SCIC *	Ethifinance	France	EthiFinance	7	250
6	GES Investment Services International AB *	Ges	Sweden	GES ESG Screening	39	4000
7	Greeneye	Greeneye	Israel	Eiris	2	30
8	imug Beratungsgesellschaft für sozial-ökologische Innovationen	Imug	Germany	Eiris	7	154
9	Korea CSR Research Service	Kocsr	Korea	Eiris	3	293
10	oekom Research AG *	Oekom	Germany	oekom Corporate Rating (Inside)	29	1060
11	Vigeo SAS *	Vigeo	France	Equitics	55	2017

*Source: (*Arise, 2014) and respective transparency matrices. Own presentation. * marks the different methodologies and the transparency matrices covered by this review. # analysts is in full time equivalents, this headcount does not cover outsourced analysts. # of companies refers to how many companies are rated.

In conclusion, the quality standard Arista 3.0 is published to facilitate transparency of SR agencies. They describe their methodology in so-called transparency matrices, which are publicly available. I apply the transparency criteria from chapter 4 to the SR agencies Eiris, Ethifinance, Ges, Oekom, and Vigeo to find out whether the information disclosed by SR agencies publicly does not allow the interpretation of the SRs' results.

5.4 Presentation of the Results

This section first presents the content of the transparency matrix template regarding the transparency criteria. Therefore I use the transparency matrix template (Arise, 2012c) and assign its individual questions to the transparency criteria. Second, the contents of the individual transparency matrices are assessed and exemplary descriptions of the methodology are given.

The transparency matrix is a mix of 531 questions. Some of them are very concrete, closed questions. For example, question F29 of the matrix asks whether the global activities of the assessed companies are covered or not. Others are broad, open questions. For example, question G15 of the matrix asks for an outline of principles followed during the actual assessment and evaluation of companies. This gives the agencies the possibility to decide how much of their methodology they disclose. Consequently, the level of transparency differs among the agencies although all of them filled out the same template.

Table 9 shows which transparency criteria have an equivalent question in the transparency matrix template and what these questions ask for. The template is detailed concerning the set of companies under analysis. Further, agencies need to inform about broad areas of the phenomenon and sources of inspiration. This continues to be of interest when relevant dimensions and individual variables are asked for. No questions ask for the primary objective and the definition of corporate sustainability. Less concrete questions concern subsequent steps of the construction. In fact, question G15 simply asks for principles followed during the assessment process. This may include the derivation of weights, the normalization, and aggregation.

The closed questions can often be complemented by a voluntary comment. However, the questions rarely force the agencies to disclose in-depth information. Instead references to internal documentation seems to be enough. Especially concerning the assessment of companies, there are no questions that ask for specific details of the methodology.

Table 9: Transparency criteria and assigned questions of the transparency matrix Arista 3.0.

#	Element	Corresponding questions and explanation	
Theoretical Framework			
TC1	Definition of the primary objective	–	
	Objects of analysis		
TC2	Exclusion criteria	E16, L, L31	E16 asks if exclusion criteria are used at all. Sheet L describes exclusion criteria as a separate methodology. L31 asks if decisions on exclusions are left to customers.
TC3	Set of companies	D40, F24-28, F62	D40 asks for explanation of industry-specific approaches. F24-28 ask how companies are selected. F62 asks for clarification on pre-selection of companies.
TC4	Scope of individual company	F29-31, F63	F29-31 ask which parts of a company are considered. F63 asks for clarification of the scope.
	Phenomenon		
TC5	Def. of corporate sustainability	–	
TC6	Global set of dimensions	D02-08, D39	D02-08 ask whether broad fields are included, e.g. ESG data. D39 asks for clarification of conceptual underpinnings.
Data selection			
TC7	Selection criterion	–	
TC8	Set of relevant dimensions	E01-02, E04-14 E29-30	E01-02 ask for the number of dimensions and sub-dimensions. E04-14 ask how many variables belong to one of eleven given dimensions. E29-30 asks for dimensions covered by the framework.
TC9	Selected variables	E03, E04-14, E18-28	E03 asks for the number of variables considered. E04-14 ask how many variables are considered in each of eleven given dimensions. E18-28 ask for variables covered by the framework.
TC10	Scaling variable	–	
Derivation of Weights			
TC11	Method to derive weights	G15	G15 asks for principles followed during the assessment process.
TC11a	Participants	G07-11 G16	G07-G11 ask which groups of participants are involved in the overall assessment process, i.e. not specifically about weighting. G16 ask for clarification on the role of an external committee.
TC11b	Instructions	–	
TC12	Set of weights	–	
Normalization			
TC13	Method for normalization	G15	G15 asks for principles followed during the assessment process.
TC13a	Reference values	–	
Aggregation			
TC14	Aggregation rule	G15	G15 asks for principles followed during the assessment process.

Source: (Arise, 2012c). Own presentation. 3rd column correspond to questions in the transparency matrix.

Table 10: Elements disclosed by SR agencies in their transparency matrices.

#	Element	Eiris			Ethifinance			Ges			Oekom			Vigeo		
		R	**D**	**J**	**R**	**D**	**J**	**R**	**D**	**J**	**R**	**D**	**J**	**R**	**D**	**J**
Theoretical Framework																
TC1	Def. of pri. objective	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
	Objects of analysis															
TC2	Exclusion criteria	☑[1]	☐[1]	☐[1]	☑[2]	☑	☐	☑[1]	☐[1]	☐[1]	☑	☑	☐	☑[1]	☐[1]	☐[1]
TC3	Set of companies	☑	☑	☐	☑[1]	☐[1]	☐[1]	☑	☑	☐	☑	☑	☐	☑	☐	☐
TC4	Scope indiv. Comp.	☑	☑	☐	☑	☑	☐	☑	☑	☐	☑	☑	☐	☑	☑	☐
	Phenomenon															
TC5	Def. of corp. sust.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☑	☐	☐	☐	☐	☐
TC6	Global set of dim.	☑[3]	☐	☐	☑[3]	☐	☐	☑[3]	☐	☐	☑[3]	☐	☐	☑[3]	☑	☐
Data selection																
TC7	Selection criterion	☐	☐	☐	☐	☐	☐	☐	☐	☐	☑	☐	☐	☐	☐	☐
TC8	Set of rel. dimensions	☑	☑	☐	☑	☑	☐	☑	☑	☐	☑	☑	☐	☑	☑	☐
TC9	Selected variables	☑	☐	☐	☑	(☑)[4]	☐	☑	(☑)[4]	☐	☑	☑	☐	☑	☐	☐
TC10	Scaling variable	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Derivation of Weights																
TC11	Method to d. weights	☐[5]	☐[5]	☐[5]	☐[5]	☐[5]	☐[5]	☐	☐	☐	☑	☐	☐	☐[5]	☐[5]	☐[5]
TC11a	Participants	☑	☑[5]	☐[5]	☑	☑[5]	☐[5]	☐	☐	☐	☑	☑	☐	☐[5]	☐[5]	☐[5]
TC11b	Instructions	☐[5]	☐[5]	☐[5]	☐[5]	☐[5]	☐[5]	☐	☐	☐	☐	☐	☐	☐[5]	☐[5]	☐[5]
TC12	Set of weights	☑[5]	☐[5]	☐[5]	☑[5]	☐[5]	☐[5]	☐	☐	☐	☑	☐	☐	☑[5]	☐[5]	☐[5]
Normalization																
TC13	Method for normaliz.	☑	☑	☐	☑	☑	☐	☐	☐	☐	☑	☑	☐	☑	☑	☐
TC13a	Reference values	☑	☐	☐	☑	☐	☐	☐	☐	☐	☑	☐	☐	☑	☐	☐
Aggregation																
TC14	Aggregation rule	☐[5]	☐[5]	☐[5]	☐[5]	☐[5]	☐[5]	☐	☐	☐	☐	☐	☐	☐[5]	☐[5]	☐[5]

Sources: (Eiris, 2012; EthiFinance, 2012; Ges, 2012; Oekom, 2012; Vigeo, 2012). An elaborate list of references is included in appendix 4 to document each box. Own presentation. R, D, and J refer to the three transparency levels reference, description, and justification. ☑ refers to "Yes", ☐ refers to "No".
Remarks: [1] decision is left to the customer. [2] no exclusion criteria used. [3] list of international conventions. [4] partial examples only. [5] only applicable to sub-phenomena, as final aggregation of indicators is left to the customer.

The individual transparency matrices for Eiris, Ethifinance, Ges, Oekom, and Vigeo are analyzed next. This analysis shows whether they make use of the possibility to disclose the construction of their SRs. Table 10 gives an overview of the contents covered by the agencies' transparency matrices. The following paragraphs describe the disclosure concerning the five steps. A list of citations for every criterion and every agency is included in appendix 4.

The review of the five transparency matrices confirms that no agency describes a primary objective. In the description of their company, some agencies state the overall objectives of their company. For example, Eiris (2012) writes: “We work to help our customers develop the market in ways that benefit investors, asset managers, and the wider world” (q. A41). Clearly, such statements are not precise enough to be applied as a primary objective during the construction of an SR. Other agencies refer to themselves as pioneers and praise the transparency their products enable for investors.

The objects of analysis are well documented. Oekom (2012, q. E18) lists exclusion criteria. Eiris, Ethifinance, and Ges use exclusion criteria, but leave it to the customers to decide which ones they want to apply. Closed questions ask for characteristics used to identify the set of companies. In open questions, three agencies specify the stock indices they cover. While all agencies claim industry-specific approaches, the set of companies is not differentiated accordingly. The scope of the individual company is described by all agencies. For example, Eiris (2012, q. F63) explicitly names equity thresholds for subsidiaries and associated companies.

Concerning the phenomenon, there is no explicit definition of corporate sustainability by any company. Oekom (2012, q. D39) refers to the definition by the Frankfurt-Hohenheim Guidelines (Balz et al., 2000). All agencies refer to conventions like the UN Global Compact, the Global Reporting Initiative, the UN Declaration on Human Rights, etc. as inspiration for their SR. Eiris (2012, q. D39) lists more than 40 such conventions. However, Vigeo is the only one explicitly describing an exhaustive global set of dimensions that categorizes the conventions.

Oekom (2012) is the only one referring to a selection criterion by stating, “when the selection was made, special emphasis was placed on the functional context, relevance, data availability and data quality” (q. D39). However, Oekom misses to specify what ‘functional context’ and ‘relevance’ it considers material. The template explicitly asks to state which out of eleven given dimensions are included. Thus, all agencies describe the set of relevant dimensions. In addition, Eiris, Ethifinance, Oekom, and Vigeo list an exhaustive list of (sub-)dimensions. For example, Vigeo (2012, q. D39) states that they consider the six dimensions human resources, human rights at the workplace, environment, business behavior, corporate governance, and community involvement. Another example, Oekom (2012, q. E30) details two dimensions. Its environmental dimension includes the sub-dimensions environmental management, products and services, and eco-efficiency. Its social dimension includes staff and suppliers, society and product responsibility, and corporate governance and business ethics. Ges only names few

exemplary sub-dimensions and variables. Vigeo and Eiris do not name any variables, instead they list sub-dimensions in the corresponding fields. Oekom (2012, q. E18-26) is the only agency naming all 119 variables that its SR for the pulp and paper industry includes. None of the agencies refers to or describes scaling variables.

The information is less exhaustive for the remaining parts of the SR construction. Eiris, Ethifinance, Oekom, and Vigeo refer to weights at different parts of their transparency matrices. Oekom (2012, q. G04) and Ges (2012, q. G04) create a one-dimensional SR based on their own weights. The other agencies only apply weights to aggregate sub-phenomena that may be weighted by customers. Oekom (2012) refers to their method to derive weights as follows: “On the basis of a scientifically developed model, the weighting of the individual sections of the rating are adapted for each industry, depending on the specific risk profile for each industry” (q. D40). Oekom does not elaborate on the ‘scientifically developed model’ nor the risk criteria it uses. Later, Oekom (2012, q. E35) describes weights of the social and environmental dimension, but not for the sub-dimensions it described earlier. Thus, the set of weights remains unclear. Ges (2012) states that “all assessments must follow the relevant manual” (q. G15) and does not bring forward any more details about its assessment of companies. Ges does not describe or refer to weights, normalization, or an aggregation rule. The referred manual is not publicly available. Concerning participants, Eiris, Ethifinance, and Oekom describe an external board of experts that is consulted concerning the methodology including the derivation of weights. However, they do not refer or describe instructions.

Eiris, Ethifinance, Oekom, and Vigeo describe a very similar description of the normalization method. All of them use grading for normalization, also due to the predominant use of qualitative variables. For example, Vigeo (2012, q. G12) states: “The scores that are attached to specific performances or characteristics of a company are indicated by the 'scoring sentences' in the database.” This example is representative of all the four agencies. All of them refer to reference values in a manual that is not publicly available. Oekom (2012, G15) vaguely refers to its best-in-class grading approach, where reference values are industry-specific and depend on a minimum grade. This implies an element of absolute evaluation, rather than a purely relative evaluation. None of the matrices allows conclusions about the absolute level of requirements a company has to fulfill in order to be considered sustainable.

Lastly, none of the transparency matrices refers to an aggregation rule although one-dimensional results are implied at various places. One may assume that the weighted arithmetic mean is perceived to be the only way to aggregate indicators.

5.5 Discussion of the Results

There are two conclusions that can be drawn from these results. First, the transparency matrices do not provide sufficient information to interpret the SRs. Second, the objectives of the standard Arista 3.0 are missed concerning the construction of SRs.

5.5.1 Level of transparency

The first conclusion is based on two findings. First, there are major parts of the construction of SRs not referred to, let alone described. Second, references outweigh descriptions.

Two major parts that are not mentioned are the primary objective of the SR and the definition of corporate sustainability. In other words, the transparency matrices do not inform the reader about the purpose of the SRs, their intentions, and intended use. Further, they do not define what they measure in theory. The reader has to assume that broad dimensions define the phenomenon and the reader has to suspect their operationalization. Chapter 3 illustrates that a primary objective and the definition of the phenomenon are at least implicitly necessary to construct an SR. In reverse, they are also necessary pieces of information for the interpretation of the SR. Furthermore, it is difficult to identify an SR matching a customer's objectives without being informed about the underlying objective of the SR.

Other major parts that are not mentioned concern the derivation of weights, the reference values of grading, and the aggregation rule. The information about weights is so sparse, that it remains unclear whether a systematic approach to the derivation of weights exists at all. Concerning the normalization of weights, clear references imply systematic grading approaches. However, when qualitative parts are a dominating part of the SR, these grading approaches change the whole SR. Does the SR incorporate ambitious requirements for each variable or is it a collection of minimum requirements? If an agency does not explain its grading approach for the variables, then the results cannot be interpreted.

The second finding supporting the first conclusion is the predominance of references without descriptions. The numerous references to elements of the SR confirm their importance for the

construction of an SR. At the same time, mentioning the mere existence of weights, reference values, and selected variables does not enable any interpretation. The motivation behind these referrals remains unclear. If agencies are worried to disclose the concrete weights, they should at least disclose the methods they use to derive weights. The same applies to reference values during grading. Here, agencies need to describe the methods and considerations that lead to the reference values. This often necessitates definitions, as references to relevance or risks do not disclose what kind of relevance and risk is implied.

5.5.2 Objectives of Arista 3.0

The second conclusion is that the objectives of Arista 3.0 are missed concerning the construction of SRs. This is also based on two findings. First, based on the other conclusion, transparency objectives are not reached. Second, the different content of the transparency matrices questions Arista 3.0's accreditation process.

Arista 3.0 has the goal "to avoid or make transparent 'black box' approaches where it is unclear how the assessment/rating result is achieved" (Arise, 2012a, p. 5). This is not the case. Especially the template's questions about the SR assessment are too broad and do not tackle major parts of the SR's construction. In addition, the lack of detailed answers to open questions by agencies determines black box approaches further. Question G15 is crucial in this context. It asks for principles being followed during the assessment. This does not require the disclosure of how the grading is actually achieved or how weights are derived. The weak coverage of these areas of enhanced by some agencies who hardly answer to this question at all.

The different level of information among transparency matrices is surprising considering the role of the transparency matrices in the accreditation process of Arista 3.0. The external auditor validates the contents of the individual transparency matrix. If the level of information is different between these matrices, the accreditation is also certifying different parts of an SR agency. Furthermore, the "disclose or explain" policy set by the standard for non-disclosure to the public is completely ignored. There is no observable reaction to this by the external auditor who verifies "each and every line" (Arise, 2012b, p. 2) of the transparency matrix.

In the words of the hypothesis, the empirical review of this chapter has confirmed that the transparency matrices do not enable the interpretation of SRs. The selected variables are not described with sufficient detail. Furthermore, elementary parts of the construction are not described at all. The transparency matrices do not describe neither which weights are applied, or

how they are derived, or how data is graded, or which aggregation rule is used. Furthermore, the SR agencies do not describe primary objectives, i.e. the purpose of their SR, and do not define the phenomenon they measure. With respect to the construction of SRs, Arista 3.0 does not ensure sufficient transparency to enable the interpretation of the SRs.

5.6 Remarks

In addition to the confirmation of the hypothesis, there is another finding to be noted. The chapter demonstrates that the transparency criteria are useful to find out whether a publication contains sufficient information to enable the interpretation of an SR. The same analysis could be carried out for documents handed out to customers and companies. A non-systematic review of public information on the agencies' websites suggests results very similar to the analysis of the transparency matrices.

Although this thesis is strictly concerned about transparency, some further remarks concerning the observed methodology and strategy of SR agencies conclude the empirical findings.

The reasons for major differences among SRs remain unclear. Table 11 shows an overview of the number of indicators considered by each SR agency and the share that is partly or fully based on qualitative information.

Table 11: Indicators considered by SR agencies.

	Eiris	**Ethifinance**	**Ges**	**Oekom**	**Vigeo**
Number of indicators	338	190	65	119	305
Number of indicators with qualitative element [1]	277	104	43	62	250
Share of indicators with qualitative element	82%	55%	66%	52%	82%

Sources: (Eiris, 2012; EthiFinance, 2012; Ges, 2012; Oekom, 2012; Vigeo, 2012). Own presentation. This data is self-reported by the SR agencies. Definitions of indicators may vary.
Remarks: [1] calculated values

If we assume that the self-reported numbers are roughly comparable, then table 11 illustrates two things. First, more than half of the indicators are influenced by qualitative variables. Second, the number of indicators deemed necessary to evaluate corporate sustainability varies.

The first finding hints at the lack of comparable quantitative information. Data availability is limited concerning the supply chain and companies' see their degree of vertical integration as strategic information. Therefore, SR agencies need to resort to qualitative data. Qualitative

data brings along the necessity of grading and the inherent subjectivity in many cases. This insight may be surprising to people unfamiliar with SRs, e.g. financial analysts who are used to quantitative data. Therefore, SR agencies may hide the concrete variables used to measure corporate sustainability intentionally. This speculation corresponds to the findings of Chelli and Gendron (2012), who find that SR agencies promote an 'ideology of numbers' without clearly stating their objectives. It also corresponds to the suspicion that SR agencies do not want to publish their primary objective in order to serve multiple customers.

Concerning the finding, it is surprising to see that the number of indicators differ by factor five. Eiris and Vigeo may count all indicators in their database and not only the ones used for assessment. However, one may wonder how many indicators are necessary to measure corporate sustainability. Saaty notes (1987, p. 163): "A general rule is that the hierarchy should be complex enough to capture the situation, but small and nimble enough to be sensitive for changes." A SR comprising around 100 indicators with roughly equal weights and aggregated by the weighted arithmetic mean will hardly react to drastic changes of one or two indicators.

One might describe SRs quite brutally and condensed: SRs are aggregates of a collection of criteria that is available and somehow connected to a notion of sustainability and responsibility. This collection is not based on systematic methods to select, weight, and aggregate these criteria. Even though this may be not true, the empirical findings show that the five agencies do little to prevent this impression through the industry's most ambitious transparency initiative. It seems like the SR agencies rely on the trust of their stakeholders. In other words, stakeholders need to believe that Eiris, Ethifinance, etc. are experts in the field of SRs and their results do not need to be questioned. If this is sufficient, the SR industry would not need a transparency initiative. However, this ignores the fact that there is no definition of corporate sustainability. Stakeholders cannot trust agencies to measure something that nobody can define in the 'correct' way. Thus, transparency is needed in addition to trust, given the different concepts of sustainability, contradictory opinions, and various fields of application. Every SR is different and addressees cannot rely on an agency to tell them what is sustainable or not without comprehending what the SR actually measures.

Some agencies opt out of the one-dimensional approach by offering measures for subphenomena only. The customer can decide how she wants to weight and aggregate these subphenomena. The composite indicator framework is very suitable to show that this approach does not avoid ambiguous decisions as long as variables are not comparable. The framework

and its conclusions are applicable on every level. The two exemplary primary objectives in chapter 3 illustrated their conflicting influence even on detailed decisions. For example, carbon emissions in a country without a price on carbon are not material to an SR that aims to predict future financial performance. Consequently, the sub-phenomena like climate impact or environmental performance will be biased by the primary objective. Some sustainability proponents may find it wrong, to define sustainability based on financial materiality. However, an SR is a mere product in the end. Eventually, customers buy the ones that reflect their concept of sustainability.

As discussed in chapter 4, the addressees of sufficient information to interpret an SR are debatable. However, it must be in the interest of SRs to show the public that their approach is suitable to integrate sustainability in financial markets. Therefore, they have to prove that the interpretation of their results is in line with their stakeholders' objectives, too.

6 Conclusion

The goal of this thesis was to set up transparency criteria for the construction of SRs based on the axiom that an agency needs to enable the SR's interpretation by disclosing information. For this purpose, it was necessary to find a research framework that is applicable to the analysis of the construction of SR. It was supposed to help identify elements of the construction that have an influence on the results. The composite indicator framework shares key characteristics with SRs. I demonstrated the applicability of the composite indicator framework to the analysis of SRs by assigning constituents to their respective equivalents and by presenting a formal problem setting that applies equivalently to SRs and composite indicators.

Next, I identified five steps in the construction of composite indicators that influence the results. These are the theoretical framework, the selection of data, the derivation of weights, the normalization, and the aggregation. I explained how to construct an SR based on these five steps. I pointed out how where alternative exists and which consequences methodological decisions have for the SR score and its interpretation.

These findings were used to identify roughly 14 elements that need to be disclosed to enable the interpretation of an SR. These elements represent the key finding of this thesis in the form of the transparency criteria presented by table 6. Each of them demands the description of an element that is part of the construction of an SR. While describing these elements enables the interpretation of an SR, a mere reference to these elements is not sufficient. I justified the transparency criteria based on the need to interpret SRs and I presented arguments that support the call for public disclosure. Yet, I restrained from identifying the definite addressees with justified claims.

In a last step, I presented an empirical review to demonstrate the application of the transparency criteria. The transparency matrices, which are published by five SR agencies in order to be accredited by Arista 3.0, were found to be incomplete. The disclosed information is not sufficient to interpret the SRs and that the Arista 3.0's objective to prevent 'black box' approaches is missed. The good news is that the application of the transparency criteria was able to identify the parts where disclosure needs to be increased. Among others, these are the primary objective to describe the purpose of an SR, the weights to describe which parts are considered important, and the reference values to know how qualitative data is graded.

I conclude that the transparency criteria are useful to analyze whether disclosed information is sufficient to enable the interpretation of SRs. Furthermore they point out room for improvement. I also conclude that SRs are essentially an aggregation of data and decisions made during their construction. Due to the lack of a 'correct' definition of corporate sustainability, SRs may measure very different concepts. As long as this is traceable, it does not diminish the need for SRs.

This thesis analyzed transparency concerning the construction of SRs. Three lines of research should be pursued further. They concern the construction itself, the comparison of SRs, and the transparency concerning other areas of SRs.

First, academics should explore the construction of SRs further. Chapter 3 has shown that many interdependencies and considerations exist. Although many options are justifiable, some approaches are advantageous compared to others. Partly methods need to be developed further. For example, quantitative measures are still difficult to include. Their inclusion would enable the consideration of absolute levels of environmental impacts. Another example is the derivation of weights. Here the analytical hierarchy process is a promising method that does not seem to be used by SR agencies.

Second, the comparison of SRs has been investigated by various authors (e.g. Balz et al., 2000; Delmas et al., 2013; Schäfer et al., 2006). However, these works remain superficial or focus on single variables. The comparison of dimensions, variables, weights, and grading methods has not been studied. As an SR is made up of these elements, they are necessary to question the societal role and the political support of SRs. Of course, this line of research requires access to information that the SR agencies do not disclose to the public at this time.

Lastly, an SR agency should be transparent about other things than its construction. Arista 3.0 includes various other elements. However, this thesis has shown that Arista 3.0 does not necessarily push SR agencies to be sufficiently transparent. The shortcomings of the transparency concerning the construction of SRs may indicate that there is room for improvement in other areas as well.

SRs are a powerful lever to foster sustainable development in the economy. Their construction is essential in this context and should meet the highest possible standards. To find out whether they do and how they could be improved, transparency is necessary. I hope that SR agencies proof to society their worthiness by increasing their transparency in the future.

References

Arrow, K. J. (1963). *Social Choice and Individual Values*. New York: John Wiley & Sons, Inc.

Association for Responsible Investment Services (Arise). (2012a). *Responsible Investment Research Standard Arista 3.0*. Retrieved from http://www.aristastandard.org/content_files/file/QS%202012/ARISTA30final09102012final.pdf

Association for Responsible Investment Services (Arise). (2012b). *Transparency Matrix Arista TM 3.0: Reading Guidelines*. Retrieved from http://www.aristastandard.org/content_files/file/TM%202013/ARISTATM3.0ReadingManual.pdf

Association for Responsible Investment Services (Arise). (2012c). *Transparency Matrix Arista TM 3.0: Template*. Retrieved from http://www.aristastandard.org/content_files/file/TM%202012/ARISTATM3.010102012final.xlsx

Balz, B.-C., Döpfner, C., Forthmann, K. R., Grieble, P., Hoffmann, J., Lücker, C. F., … Schneider, H.-A. (2000). *Frankfurt-Hohenheimer Leitfaden*. Retrieved from http://www.cric-online.org/images/individual_upload/div_infos/fhl-d-05.pdf

Bassen, A., & Kovács, A. M. (2008). Environmental, social and governance key performance indicators from a capital market perspective. *Zeitschrift Für Wirtschafts- Und Unternehmensethik*, *9*(2), 182–192.

Baumann, H., & Tillman, A.-M. (2004). *The Hitch Hiker's Guide to LCA. An orientation in life cycle assessment methodology and application*. doi:10.1065/lca2006.02.008

Böhringer, C., & Jochem, P. E. P. (2007). Measuring the immeasurable — A survey of sustainability indices. *Ecological Economics*, *63*(1), 1–8. doi:10.1016/j.ecolecon.2007.03.008

Booysen, F. (2002). An overview and evaluation of composite indices of development. *Social Indicators Research*, *59*(2), 115–151. doi:10.1023/A:1016275505152

Borucke, M., Moore, D., Cranston, G., Gracey, K., Iha, K., Larson, J., … Galli, A. (2013). Accounting for demand and supply of the biosphere's regenerative capacity: The National Footprint Accounts' underlying methodology and framework. *Ecological Indicators*, *24*, 518–533. doi:10.1016/j.ecolind.2012.08.005

Bowker, G. C., & Star, S. L. (2000). *Sorting Things Out: Classification and Its Consequences*. Cambridge: MIT Press.

Chelli, M., & Gendron, Y. (2012). Sustainability Ratings and the Disciplinary Power of the Ideology of Numbers. *Journal of Business Ethics*, *112*(2), 187–203. doi:10.1007/s10551-012-1252-3

Cho, C. H., & Patten, D. M. (2013). Green accounting: Reflections from a CSR and environmental disclosure perspective. *Critical Perspectives on Accounting*, *24*(6), 443–447. doi:10.1016/j.cpa.2013.04.003

Clark, G. L., Feiner, A., & Viehs, M. (2014). *From the stockholder to the stakeholder - How sustainability can drive financial outperformance*. Retrieved from http://www.smithschool.ox.ac.uk/research/library/SSEE_Arabesque_Paper_16Sept14.pdf

Delmas, M., & Blass, V. D. (2010). Measuring corporate environmental performance: the trade-offs of sustainability ratings. *Business Strategy and the Environment*, *19*(4), 245–260.

Delmas, M., Etzion, D., & Nairn-Birch, N. (2013). Triangulating environmental performance: What do environmental ratings really capture? *Academy of Management Perspectives*, *27*(3), 255–267. doi:10.5465/amp.2012.0123

Deutsche Bank Climate Change Advisors. (2012). *Sustainable Investing - Establishing long-term value and performance*.

Dillenburg, S., Greene, T., & Erekson, H. (2003). Approaching Socially Responsible Investment with a Comprehensive Ratings Scheme: Total Social Impact. *Journal of Business Ethics*, (43), 167–177.

Döpfner, C., & Schneider, H.-A. (2012). *Nachhaltigkeitsratings auf dem Prüfstand*. Retrieved from www.cric-online.org/images/individual_upload/publikationen/nachhaltigkeitsstudie 2012.pdf

Ebert, U., & Welsch, H. (2004). Meaningful environmental indices: a social choice approach. *Journal of Environmental Economics and Management*, *47*(2), 270–283. doi:10.1016/j.jeem.2003.09.001

Esty, D. C., Levy, M., Srebotnjak, T., & de Sherbinin, A. (2005). *2005 Environmental Sustainability Index: Benchmarking national environmental stewardship*. *Environmental Law*. New Haven: Yale Center for Environmental Law, Yale University and Policy and Center for International Earth Science Information Network, Columbia University. Retrieved from http://www.yale.edu/esi

Ethical Investment Research Services Ltd (Eiris). (2012). *Transparency Matrix ARISTA TM 3.0*. Retrieved from http://www.aristastandard.org/content_files/file/TM%202013/ EIRISARISTATM3020131010.pdf

EthiFinance. (2012). *Transparency Matrix ARISTA TM 3.0*. Retrieved from http://www.aristastandard.org/content_files/file/TM%202013/EthiFinanceARISTATM3. 02013.pdf

European Commission Joint Research Centre. (2010). *International Reference Life Cycle Data System Handbook: Framework and requirements for Life Cycle Impact assessment models and indicators* (1st ed.). Luxembourg. doi:10.2788/38719

European Federation of Financial Analysts Societies, & Society of Investment Professionals in Germany. (2010). *Key Performance Indicators for Environmental, Social & Governance Issues 3.0*. Main. Retrieved from http://www.effas-esg.com/wp-content/uploads/2011/07/KPIs_for_ESG_3_0_Final.pdf

Eurosif. (2014). *European SRI Study 2014*. Retrieved from http://www.eurosif.org/publication/download/european-sri-study-2014

Galli, A., Kitzes, J., Wermer, P., Wackernagel, M., Niccolucci, V., & Tiezzi, E. (2008). An exploration of the mathematics behind the ecological footprint. *International Journal of Ecodynamics*, *2*(4), 250–257. doi:10.2495/ECO-V2-N4-250-257

Gasparatos, A. (2010). Embedded value systems in sustainability assessment tools and their implications. *Journal of Environmental Management*, *91*(8), 1613–22. doi:10.1016/j.jenvman.2010.03.014

Gasparatos, A., El-Haram, M., & Horner, M. (2008). A critical review of reductionist approaches for assessing the progress towards sustainability. *Environmental Impact Assessment Review*, *28*(4-5), 286–311. doi:10.1016/j.eiar.2007.09.002

Gasparatos, A., El-Haram, M., & Horner, M. (2009). The argument against a reductionist approach for measuring sustainable development performance and the need for methodological pluralism. *Accounting Forum*, *33*(3), 245–256. doi:10.1016/j.accfor.2008.07.006

GES Investment Services International AB (Ges). (2012). *Transparency Matrix ARISTA TM 3.0*. Retrieved from http://www.aristastandard.org/content_files/file/TM 2013/GESARISTATM3.02013.pdf

Gladman, K. (2011). *Ten things to know about responsible investment and performance*. Retrieved from www3.gmiratings.com/wp-content/uploads/2012/10/GMIRatings_ 10Things_102012.pdf

Global Initiative for Sustainability Ratings. (2013). Sustainability Ratings Standard - Component 1: Principles. Retrieved October 11, 2014, from http://ratesustainability.org/ standards/principles/

Global Reporting Initiative (GRI). (2013a). G4 Development Process. Retrieved May 22, 2014, from https://www.globalreporting.org/reporting/g4/g4-developments/Pages/ default.aspx

Global Reporting Initiative (GRI). (2013b). *G4 sustainability reporting guidelines: Implementation manual*. Retrieved from https://www.globalreporting.org/ resourcelibrary/GRIG4-Part2-Implementation-Manual.pdf

Global Reporting Initiative (GRI). (2013c). *G4 sustainability reporting guidelines: Reporting principles and standard disclosures*. Retrieved from https://www.globalreporting.org/ resourcelibrary/GRIG4-Part1-Reporting-Principles-and-Standard-Disclosures.pdf

Global Reporting Initiative (GRI). (2013d). *Sustainability topics for sectors: What do stakeholders want to know?*. Retrieved from https://www.globalreporting.org/resourcelibrary/sustainability-topics.pdf

Goedkoop, M., Heijungs, R., Huijbregts, M., De Schryver, A., Struijs, J., & van Zelm, R. (2013). *ReCiPe 2008: A life cycle impact assessment method which comprises harmonised category indicators at the midpoint and the endpoint level*.

Goodall, T. (2005). Measuring corporate sustainability performance and its impact on corporate financial performance. In C. Böhringer & A. Lange (Eds.), *Applied Research in Environmental Economics*. Heidelberg: Physica-Verlag.

Hardegree, G. (2001). Measurement. In *Notes on Philosophy of Science*. Amherst: University of Massachusetts. Retrieved from http://people.umass.edu/gmhwww/382/text.htm

Hardi, P., & Zdan, T. (1997). *Assessing sustainable development: Principles in practice*. Retrieved from http://www.iisd.org/pdf/bellagio.pdf

Hawken, P. (2004). *Socially responsible investing*. Retrieved from http://q.investorideas.com/CO/Marion_Institute/SRIReport.pdf

Hermann, B. G., Kroeze, C., & Jawjit, W. (2007). Assessing environmental performance by combining life cycle assessment, multi-criteria analysis and environmental performance indicators. *Journal of Cleaner Production*, *15*(18), 1787–1796. doi:10.1016/j.jclepro.2006.04.004

Hsieh, N. (2008). Incommensurable Values. (E. N. Zalta, Ed.) *The Stanford Encyclopedia of Philosophy*. Retrieved from http://plato.stanford.edu/archives/fall2008/entries/value-incommensurable/

Hsu, A., Johnson, L., & Lloyd, A. (2013). Measuring progress: A practical guide from the developer of the Environmental Performance Index (EPI). New Haven: Yale Center for Environmental Law & Policy.

Ilinitch, A. Y., Soderstrom, N. S., & E. Thomas, T. (1998). Measuring corporate environmental performance. *Journal of Accounting and Public Policy*, *17*(4-5), 383–408. doi:10.1016/S0278-4254(98)10012-1

Koellner, T., Weber, O., Fenchel, M., & Scholz, R. (2005). Principles for sustainability rating of investment funds. *Business Strategy and the Environment*, *14*(1), 54–70. doi:10.1002/bse.423

Krajnc, D., & Glavic, P. (2005). A model for integrated assessment of sustainable development. *Resources, Conservation and Recycling*, *43*, 189–208. Retrieved from http://www.sciencedirect.com/science/article/B6VDX-4DD9CND-1/2/a688e50572ae3cd75f4c3325d1a06a1b

Krajnc, D., & Glavič, P. (2005). How to compare companies on relevant dimensions of sustainability. *Ecological Economics*, *55*, 551–563. doi:10.1016/j.ecolecon.2004.12.011

Mackenzie, C., Rees, W., & Rodionova, T. (2013). Do Responsible Investment Indices Improve Corporate Social Responsibility? FTSE4Good's Impact on Environmental Management. *Corporate Governance: An International Review*, *21*, 495–512. doi:10.1111/corg.12039

Margni, M., & Curran, M. A. (2012). Life Cycle Impact Assessment. In *Life Cycle Assessment Handbook* (pp. 67–103). John Wiley & Sons, Inc. doi:10.1002/9781118528372.ch4

Moldan, B., Hák, T., Kovanda, J., Havranek, M., & Kusková, P. (2004). OECD World Forum on Key Indicators. In *Composite Indicators of Environmental Sustainability*. Palermo: OECD. Retrieved from http://www.oecd.org/oecdworldforum

Moldan, B., Janoušková, S., & Hák, T. (2012). How to understand and measure environmental sustainability: Indicators and targets. *Ecological Indicators*, *17*, 4–13. doi:10.1016/j.ecolind.2011.04.033

Munda, G. (2005). Multiple Criteria Decision Analysis and Sustainable Development. In J. Figueira, S. Greco, & M. Ehrgott (Eds.), *Multiple Criteria Decision Analysis: State of the Art Surveys* (Vol. 78, pp. 953–986). New York: Springer. doi:10.1007/b100605

Munda, G. (2008). *Social multi-criteria evaluation: Methodological foundations and operational consequences*. *European Journal of Operational Research* (Vol. 158). Berlin, Heidelberg: Springer-Verlag.

Munda, G. (2012). Choosing Aggregation Rules for Composite Indicators. *Social Indicators Research*. doi:10.1007/s11205-011-9911-9

Munda, G., & Nardo, M. (2005). Constructing Consistent Composite Indicators : the Issue of Weights. Luxembourg: Office for Official Publications of the European Communities.

Munda, G., & Nardo, M. (2009). Noncompensatory/nonlinear composite indicators for ranking countries: a defensible setting. *Applied Economics*. doi:10.1080/00036840601019364

Nardo, M., Saisana, M., Saltelli, A., & Tarantola, S. (2005). *Tools for composite indicators building*. *Analysis* (p. 134). Ispra. doi:10.1038/nrm1524

Novethic. (2013). *Overview of ESG rating agencies*. Retrieved from http://www.novethic.fr/novethic/upload/etudes/2013_overview_ESG_rating_agencies.pdf

O'Neill, J. (1993). *Ecology, policy and politics*. New York: Routledge.

OECD. (2008). *Handbook on Constructing Composite Indicators: Methodology and User Guide*. *Methodology* (Vol. 3, p. 162). doi:10.1787/9789264043466-en

Oekom. (2012). *Transparency Matrix ARISTA TM 3.0*. Retrieved from http://www.arista standard.org/content_files/file/TM 2013/OekomARISTATM3.02013.pdf

Ott, W. R. (1978). *Environmental indices: Theory and practice*. Ann Arbor: Ann Arbor Science Publisher.

Parris, T. M., & Kates, R. W. (2003). Characterizing and measuring sustainable development. *Annual Review of Environment and Resources*, *28*(1), 559–586. doi:10.1146/annurev.energy.28.050302.105551

Pearce, D. W., & Turner, R. K. (1990). *Economics of natural resources and the environment*. Baltimore: The Johns Hopkins University Press.

Ramos, T. B., & Caeiro, S. (2010). Meta-performance evaluation of sustainability indicators. *Ecological Indicators*, *10*(2), 157–166. doi:10.1016/j.ecolind.2009.04.008

Robecosam. (2013). *Measuring intangibles: Robecosam's corporate sustainability assessment methodology*. Retrieved from http://www.robecosam.com/images/CSA_methodology_us.pdf

Rose, A., van Ast, L., Bolster, P., Knight, C., & Bennett, L. (2013). *Environmental disclosures*. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/298793/Disclosures_summary_aaf775.pdf

Saaty, R. W. (1987). The analytic hierarchy process: what it is and how it is used. *Mathematical Modelling*, *9*(3-5), 161–176.

Sadowski, M., Whitaker, K., & Ayars, A. (2011). *Rate the raters phase 3: Uncovering best practices*. Retrieved from http://www.sustainability.com/library/rate-the-raters-phase-three#.UFjA36TybIs

Sadowski, M., Whitaker, K., & Buckingham, F. (2010). *Rate the raters phase 2: Taking inventory of the ratings universe*. Retrieved from http://www.sustainability.com/library/rate-the-raters-phase-two

Sadowski, M., Whitaker, K., Lee, M., & Ayars, A. (2011). *Rate the raters phase 4: The necessary future of ratings*. Retrieved from http://www.sustainability.com/library/rate-the-raters-phase-four#.UFjAu6TybIs

Scalet, B. M., Garcia Muñoz, M., Sissa, A. Q., Roudier, S., & Delgade Sancho, L. (2013). *Best Available Techniques (BAT) reference document for the manufacture of glass*. Seville: Joint Research Centre of the European Commission. doi:10.2791/69502

Schäfer, H., Beer, J., Zenker, J., & Fernandes, P. (2006). *Who is who in corporate social responsibility rating ?*. Gütersloh. Retrieved from http://www.bertelsmann-stiftung.de/cps/rde/xbcr/SID-0A000F14-FF792680/bst/xcms_bst_dms_18175_18176_2.pdf

Schmidheiny, S. (1996). Eco-efficiency and sustainable development. *Risk Management*, *43*(7).

Sen, A. K. (1982). *Choice, Welfare And Measurement*. Oxford: Basil Blackwell.

Sharpe, A. (2004). *Literature review of frameworks for macro-indicators*. Retrieved from http://www.csls.ca/reports/LitRevMacro-indicators.pdf

Stevens, S. S. (1946). On the Theory of Scales of Measurement. *Science (New York, N.Y.)*, *103*(2684), 677–80. doi:10.1126/science.103.2684.677

Sustainability Accounting Standards Board. (2013). *Pharmaceuticals Sustainability Accounting Standard*. Retrieved from http://www.sasb.org/standards/download/health-care/

The International Integrated Reporting Council. (2013). *The international IR framework*. Retrieved from http://www.theiirc.org/international-ir-framework/

Tyteca, D. (1996). On the measurement of the environmental performance of firms - A literature review and a productive efficiency perspective. *Journal of Environmental Management*, *46*(3), 281–308. doi:10.1006/jema.1996.0022

United Nations (UN). (1987). *Report of the World Commission on Environment and Development: Our Common Future (The Brundtland Report)*. doi:10.1080/07488008808408783

United Nations (UN). (1992). *Rio Declaration on Environment and Development*. Retrieved from http://www.un.org/documents/ga/conf151/aconf15126-1annex1.htm

United Nations (UN), European Commission (EC), Food and Agriculture Organization of the United Nations (FAO), International Monetary Fund (IMF), Organisation for Economic Co-operation and Development (OECD), & The World Bank (World Bank). (2014). *System of Environmental-Economic Accounting 2012: Central Framework*. New York: United Nations.

United Nations Development Programme. (2014). *Human Development Report 2014: Technical notes*. Retrieved from http://hdr.undp.org/sites/default/files/hdr14_technical_notes.pdf

United Nations Global Compact, & Business for Social Responsibility. (2012). *Nachhaltigkeit in der Lieferkette*. Retrieved from http://www.globalcompact.de/sites/default/files/jahr/publikation/nachhaltigkeit_in_lieferkette_120818.pdf

United Nations Principles for Responsible Investment (UNPRI). (2013). *Integrated analysis: How investors are addressing environmental, social and governance factors in fundamental equity valuation*. Retrieved from http://www.unpri.org/viewer/?file=wp-content/uploads/Integrated_Analysis_2013.pdf

United Nations Principles for Responsible Investment (UNPRI). (2014). PRI goes from strength-to-strength as signatory assets top US$ 45 trillion. Retrieved August 01, 2014, from http://www.unpri.org/whatsnew/pri-goes-from-strength-to-strength-as-signatory-assets-top-us-45-trillion/

Vigeo. (2012). *Transparency Matrix ARISTA TM 3.0*. Retrieved from http://www.aristastandard.org/content_files/file/TM 2013/VigeoARISTATM302015.pdf

Weber, M., Eisenführ, F., & von Winterfeldt, D. (1988). The effects of splitting attributes on weights in multiattribute utility measurement. *Management Science*, *34*(4), 431–445. doi:10.1287/mnsc.34.4.431

Weber, O., Koellner, T., Habegger, D., Steffensen, H., & Ohnemus, P. (2005). *The relation between the sustainability performance and the financial performance of firms*. doi:10.1504/PIE.2008.019127

Weidema, B. P., Hischier, R., Mutel, C., Nemecek, T., Reinhard, J., Vadenbo, C. O., & Wernet, G. (2013). *Overview and methodology - Data quality guideline for the ecoinvent database* (Vol. 3). St. Gallen.

Welsch, H. (2005). Constructing meaningful sustainability indices. In C. Böhringer & A. Lange (Eds.), *Applied Research in Environmental Economics* (pp. 7–22). Heidelberg: Physica-Verlag.

Windolph, S. E. (2011). Assessing Corporate ustainability through ratings: Challenges and their causes. *Journal of Environmental Sustainability*, *1*(1), 1–22. doi:10.14448/jes.01.0005

World Business Council for Sustainable Development (WBCSD), & World Resources Institute (WRI). (2004). *The Greenhouse Gas Protocol: A Corporate Accounting and Reporting Standard - Revised Edition*. Retrieved from www.ghgprotocol.org

Wrisberg, N., & de Haes, H. A. (2002). *Analytical tools for environmental design and management in a systems perspective: The combined use of analytical tools*. Dordrecht: Kluwer Academic Publishers.

Zhou, P., Ang, B. W., & Poh, K. L. (2006). Comparing aggregating methods for constructing the composite environmental index: An objective measure. *Ecological Economics*, *59*(3), 305–311. doi:10.1016/j.ecolecon.2005.10.018

Appendix 1: Global Sets of Environmental Dimensions

Sustainability ratings have to define the phenomenon they measure. Corporate environmental performance is usually a sub-phenomenon. I analyzed three different approaches to find out whether a global set of dimensions can be identified for the environment. Statistical offices, reporting standard setters, and life cycle analysts have to solve the same problem. Statistical offices analyze nations' interactions with the environment, reporting standards structure corporate environmental reporting, and life cycle analysis looks at processes that impact the environment. The following sections present the three different sets shortly.

Statistical offices

Representatives from statistical offices around the world gathered in 1993 after the publication of the *Handbook of National Accounting: Integrated Environmental and Economic Accounting* to form the London Group on Environmental Accounting. The System of Environmental-Economic Accounting (SEEA) is supposed to be the "international statistical standard for environmental-economic accounting" (UN et al., 2014, p. 3). To classify interactions with the environment, input and output flows are differentiated.

By definition, material input flows that affect the environment do so by depleting natural resources. The classification of natural inputs is based on environmental assets. They encompass six classes of resources and land as listed by table A1.1. Land accounts for the provision of space. It cannot be depleted materially, but is limited, too (UN et al., 2014, p. 134).

Outputs of the economy are residuals such as emissions and return flows of water. Their cate-

Table A1.1: Set of dimensions for environmental impacts by statistical offices

Input related: "Natural inputs"	**Output related: "Residuals"**
Mineral and energy resources	Ambient air and climate
Land	Wastewater
Soil resources	Waste
Timber resources	Soil, groundwater, and surface water
Aquatic resources	Noise and vibration
Other biological resources	Biodiversity and landscapes
Water resources	Radiation

Source: (UN et al., 2014, pp. 45, 99). Own presentation.

gorization is based on the domains, which receive these dimensions. The decision criterion is based on physical differences (United Nations (UN) et al., 2014, p. 13).

The SEEA classifications constitute a systematic approach to find a set of dimensions for interaction with the environment. The use of spelled out division criteria and the resulting classifications could be used be SR agencies to conceptualize environmental performance. Construction could proceed identifying the most relevant dimensions and finding variables accordingly.

Sustainability reporting standards

The GRI publishes reporting guidelines for sustainability reporting since 2000. Its fourth generation, the G4, was published in 2013. Its inclusion of the four biggest audit firms and "hundreds" of stakeholders (GRI, 2013c), public comment periods, and the employment of G3 by many of the biggest companies worldwide make it the quasi-standard within sustainability reporting (Krajnc & Glavič, 2005, p. 189; O. Weber et al., 2005, p. 7).

Its reporting guidelines on the environmental dimension aim to cover "the organization's impact on living and non-living natural systems, including land, air, water and ecosystems" (GRI, 2013b, p. 52). For its standard disclosures, GRI differentiates input and output variables, just like the statistical offices, as well as additional aspects presented in table A1.2.

In contrast to the statistical offices, the delineation of GRI's classification does not follow a single division criterion. Instead, the classification is the result of a participative process and a subsequent ordering. It represents a bottom-up approach, as described above.

Table A1.2: Set of dimensions for environmental aspects by the Global Reporting Initiative

Input related	**Output related**	**Other aspects**
Material	Emission	Biodiversity
Energy	Effluents and Waste	Products and Services
Water		Compliance
		Transport
		Overall
		Supplier Environmental Assessment
		Environmental Grievance Mechanisms

Source: (GRI, 2013c, pp. 44, 52ff.) Own presentation.

G4 distinguishes between topics and aspects. Topics are all sustainability subjects while aspects are a pre-selected set of topics considered by the guidelines. This pre-selection bases on a multi-stakeholder development process. It is not based on a precise definition of the phenomenon. Topics related to corporate sustainability were researched by a coordinating team and submitted by anyone interested. This resulted in 1612 unique topics (GRI, 2013d). All of them were assigned to the aspects from table A1.2. Consequently, the list is mutually exclusive for all aspects by definition, but not for all topics. The scope is defined on the way by the width of topics proposed. It would be a coincidence, if the stakeholder's individual proposals add up to a concept of sustainability, which equals the theoretical concept of a phenomenon, as set by an SR agency. This is due to the primary objective of GRI. As a reporting standard, it aims to maximize the inclusion of aspects and (diverging) stakeholder interests. It is not aiming to represent a single phenomenon.

Subsequently, the classification above is less systematic than the one from the statistical offices, with seven environmental aspects overlapping with the input- and output-related aspects. The classification of aspects covered by G4 does not define a global set of dimensions for environmental sustainability, but it is a collection of topics in its context. With respect to SRs, the list can serve as an inspiration.

Dimensions in product lifecycle impact analysis

Last, I shortly introduce a classification used by product lifecycle analysts. Product lifecycle analysis is arguably one of the most established approaches to assess environmental and other impacts in the business sphere. Its origins date back to 1991 with the German *Ecoscarcity* approach (Margni & Curran, 2012, p. 78). ISO 14044 standardized principles of lifecycle assessments and the Joint Research Center by the European Commission (EC-JRC) further aligned methodologies by publishing a series of handbooks (EC-JRC, 2010). The Swiss not-for-profit association *ecoinvent* offers lifecycle inventory data since 2003, with the third version of its database launched in 2014 (Weidema et al., 2013, p. 11ff.). The objects of analysis are individual transformation activities and market activities. To derive a product's footprint, all connected activities are aggregated.

The set of elementary flows that are part of the life cycle inventory are individual for each product (Margni & Curran, 2012, p. 69). In order to structure them in the ecoinvent database, they are assigned to nine so-called environmental compartments (Weidema et al., 2013, p. 63). However, the set of dimensions used for the impact assessments consists of actual im-

pacts. This step assesses the life cycle inventory by assigning each flow to one or more out of 14 impact categories, or so-called midpoints. After this 'classification' step, the 'characterization' step calculates each indicators influence on the 14 impacts. Ten out of the fourteen impact categories have a cause-effect relation to the environment.

Table A1.3: Set of dimensions for environmental impacts by life cycle analysts

Impact Categories	
Climate Change	Acidification
Ozone Depletion	Eutrophication
Respiratory Inorganics / Particulate Matter	Ecotoxicity
Ionizing Radiation	Land use
Photochemical ozone formation	Resource depletion

Source: (Margni & Curran, 2012, p. 71). Own presentation.

The impact categories do not exhaust one phenomenon exhaustively. They base on the results of a selection process of a group of scientists. Their primary objective was to identify impacts which were most pressing on a global scale, although regional differences exist (European Commission Joint Research Centre, 2010, p. 3). The resulting ten categories with respect to the environment represent a suggested set of important environmental impacts as identified by life cycle analysts on an ongoing basis. The backing of the scientific community and practitioners gives reason to believe, that this approach is a good balance between theory and feasibility. At the same time, a primary objective cannot be excluded, e.g. a European perspective on environmental problems might bias the set of dimensions. Still, the LCIA impact categories could serve as starting point, when SR agencies conceptualize environmental performance.

The three examples of classification illustrate three approaches to structure the phenomenon 'environment' theoretically before selecting the relevant dimensions. Environmental aspects of sustainability can be defined by an exhaustive set of dimensions G as illustrated by the SEEA classification of the environment. A stakeholder based, bottom-up process may lead to inapt global sets of dimensions. Lastly, the LCIA categories of impacts indicate which categories may be considered most relevant from a scientific point of view. A global sets of dimensions, whether constructed or adopted, is filtered to fit the SR's primary objective.

Appendix 2: Level of Measurement

Measurement processes have to consider the level of measurement of each variable. Four levels are typically distinguished: nominal, ordinal, interval, and ratio level. Interval and ratio data are also called metric. Nominal data is considered qualitative; the other levels are quantitative data. This classification was originally published by Stevens (2010, pp. 3, 19) and has become a matter of common knowledge in empirical studies. Based on the measurement process, when a value is assigned to an observed attribute, the level of measurement describes how a variable can be interpreted and which operations are permissible, i.e. which operations lead to meaningful results (OECD, 2008, p. 53). Thus, it differentiates between a mathematical treatment of data and the actual meaning concerning the underlying attribute. While transformations may be possible mathematically, measurement theory identifies the transformations and operations that are meaningful given the measurement process and the consequent level of measurement. Consequently, the level of measurement influences the interpretation of measures with respect to an attribute.

Each level of measurement is characterized by the set of transformations that can be applied to the variable without changing its meaning. As Stevens puts it, “In what ways can we transform its values and still have it serve all the functions previously fulfilled?” (1946). Variables are *invariant* under these transformations. Additionally, operations require variables of a certain level of measurement. Otherwise, they will not lead to meaningful results. Colloquially speaking, if operations are based on information that is not captured by the measurement process, they are impermissible. For example, the ordinal level does not permit operations based on the distance between two values because it lacks information about these intervals. As the arithmetic mean requires that information, it cannot be applied to ordinal data. The four levels of measurement are summarized in table A2.1.

The interval- and ratio-level are collectively called metric. Other levels exist in addition to these four levels of measurement. Quasi-interval variables fulfill the criteria of the ordinal scale, but may be interval variables if one assumes that the intervals between the SRs are not arbitrary. Common examples are school grades A-F and survey results on the Likert scale from ‘strongly agree = 7’ to ‘strongly disagree = 1’. Doubts are based on the presumed inability of human beings to assign grades with constant intervals. Thus, there are doubts about the consistency when survey results are interpreted as interval or ratio data.

Table A2.1: Overview of the four levels of measurement.

Level of measurement	***Set of invariant transformations F***	***Permissible operations***	***Example***
Nominal	$\boldsymbol{f: x \rightarrow p = x}$, as in any one-to-one substitution	Equivalence Mode	Qualitative categories, e.g. nationalities, gender, continents
Ordinal	$\boldsymbol{f: x \rightarrow p = f(x)}$, with $\boldsymbol{f(x)}$ being any monotonic increasing function	Rank order Median	Rankings, e.g. $\boldsymbol{a_1 > a_2 > a_3}$, best-in-class ratings, soccer league table
Interval	$\boldsymbol{f: x \rightarrow p = \alpha x + \beta, \alpha > 0, \beta \neq 0}$, any positive linear transformation	Distance Arithmetic mean	Cardinal values with arbitrary origin, e.g. temperature in °Celsius, year dates
Ratio	$\boldsymbol{f: x \rightarrow p = \alpha x, \alpha > 0}$, any homogeneous function of degree 1	Ratios Geometric mean	Cardinal values with an origin, e.g. lengths, masses, forces

Source: (Stevens, 1946; Munda, 2008; OECD, 2008). Own presentation.

Each level of measurement entails the permissible operations of the preceding levels. That is, equivalence and the mode can be determined for all levels. In order to determine ratios such as 'is twice as much as', variables need to be on the ratio level. The arithmetic mean can be calculated for variables that are measurable on the interval or ratio level.

If a variable is transformed by a function $f \in F$ with F being invariant to a lower level of measurement, the resulting variable is on the lower level of measurement of the original variable and F. An expansion of a variable measurable on the interval level by $f: x \rightarrow p = \alpha x, \alpha > 0$ will result in a variable on the interval level. This is intuitive, as there is nowhere an origin could come from. Conversely a linear transformation defined as $f: x \rightarrow p = \alpha x + \beta, \alpha > 0, \beta \neq 0$ applied to a ratio-variable will result in an interval-variable. This is intuitive as well, as adding β changes the origin. Ideally, all variables remain on the ratio level, as it enables the most operations.

These concepts apply to the construction of composite indicators as follows. Each dimension i is represented by a variable α_i. Two variables α_i and α_{i+1} can be on different levels of measurement. As variables are transformed before aggregation, their level of measurement might be reduced. More precisely, c*hanges* of the level of measurement *reduce* the amount of admissible operations. Two common aggregation rules are the arithmetic and the geometric mean. The latter is only permissible if the variables are on a ratio level. Consequently, it is advisable to maintain the level of measurement.

Appendix 3: Marginal Rate of Substitution

Appendix 3 complements section 3.6.2. It shows the calculation of the marginal rate of substitution for the weighted arithmetic mean and the weighted geometric mean.

The marginal rate of substitution is calculated for I_1 and I_2. It describes the rate at which a company j can reduce its value of indicator I_1 in exchange for an increase in indicator I_2 while maintaining the same SR score. For all calculations I assume that the sum of weights $\sum_{k=1}^{M'} w_k = 1$ and the number of indicators $M' \geq 3$.

Weighted Arithmetic Mean

The weighted arithmetic mean is defined as

$$SR_j = \sum_{k=1}^{M'} w_k I_{k,j} \text{ with } j = 1, \dots, N \text{ and } k = 1, \dots, M'.$$

This can be rewritten as

$$SR_j = w_1 I_{1,j} + w_2 I_{2,j} + \sum_{k=3}^{M'} w_k I_{k,j}\,.$$

The partial derivative of SR_j in the direction of I_1 is

$$\frac{\partial SR_j}{\partial I_{1,j}} = w_1.$$

The partial derivative of SR_j in the direction of I_2 is

$$\frac{\partial SR_j}{\partial I_{2,j}} = w_2.$$

The marginal rate of substitution is

$$MRS_{1,2} = \frac{\frac{\partial SR_j}{\partial I_{1,j}}}{\frac{\partial SR_j}{\partial I_{2,j}}} = \frac{w_1}{w_2}.$$

Weighted Geometric Mean

The weighted geometric mean is defined as

$$SR_j = \prod_{k=1}^{M'} I_{k,j}^{w_k} \text{ with } j = 1, \dots, N \text{ and } k = 1, \dots, M'.$$

This can be rewritten as

$$SR_j = I_{1,j}^{w_1} \cdot I_{2,j}^{w_2} \cdot \prod_{k=3}^{M'} I_{k,j}^{w_k}.$$

The partial derivative of SR_j in the direction of I_1 is

$$\frac{\partial SR_j}{\partial I_{1,j}} = w_1 I_{1,j}^{w_1 - 1} \cdot I_{2,j}^{w_2} \cdot \prod_{k=3}^{M'} I_{k,j}^{w_k}.$$

The partial derivative of SR_j in the direction of I_2 is

$$\frac{\partial SR_j}{\partial I_{2,j}} = I_{1,j}^{w_1} \cdot w_2 I_{2,j}^{w_2 - 1} \cdot \prod_{k=3}^{M'} I_{k,j}^{w_k}.$$

The marginal rate of substitution is

$$MRS_{1,2} = \frac{\frac{\partial SR_j}{\partial I_{1,j}}}{\frac{\partial SR_j}{\partial I_{2,j}}} = \frac{w_1 I_{1,j}^{w_1 - 1} \cdot I_{2,j}^{w_2} \cdot \prod_{k=3}^{M'} I_{k,j}^{w_k}}{I_{1,j}^{w_1} \cdot w_2 I_{2,j}^{w_2 - 1} \cdot \prod_{k=3}^{M'} I_{k,j}^{w_k}}.$$

The products can be cancelled and rearranging gives

$$MRS_{1,2} = \frac{w_1}{w_2} \cdot \frac{I_{1,j}^{w_1 - 1}}{I_{1,j}^{w_1}} \cdot \frac{I_{2,j}^{w_2}}{I_{2,j}^{w_2 - 1}}.$$

The weighted indicators can also be cancelled and now the marginal rate of substitution is

$$MRS_{1,2} = \frac{w_1}{w_2} \cdot \frac{I_2}{I_1}.$$

Appendix 4: Data Tables

The results presented in section 5.4 are based on the five reviewed transparency matrices of the SR agencies. The following tables cite the sources of the results presented in table 10.

Eiris

Table A4.1: Elements disclosed by Eiris in its transparency matrix

#	R	D	J	Source	Referenced Extracts
TC1	☐	☐	☐	–	
TC2	☑[1]	☐[1]	☐[1]	L31 D05	Decisions on exclusions are left to customers: "Yes" "The methodology is designed to be flexible for the end user/client as EIRIS recognises that no two ethical/socially responsible investors are alike."
TC3	☑	☑	☐	F24-28 F64	"EIRIS tracks and covers companies on the FTSE All Share, FTSE All World Developed, StoXX 600, S&P 350, Dax 30, MidDax, Ibex 35 and S&P500 indices"
TC4	☑	☑	☐	F29-31 F63	Global operations are covered "Yes" Subsidiaries, joint ventures and associated companies are covered "Yes" Upstream ownership (shareholding) is analyzed "No" "Company - an entity made up of a parent company, its divisions, subsidiaries, associated companies and operations." "Subsidiary - if the parent company holds more than 50% of the equity share capital in another company" "Associates - where the parent company (including its subsidiaries) holds 20-50% inclusive of the aggregate interest in the equity share capital of another company"
TC5	☐	☐	☐	–	
TC6	☑[3]	☐	☐	D02-09 D39	Methodology is inspired by: - General international codes, conventions, guidelines "Yes" - International codes on corporate sustainability "Yes" - Concepts of corporate social responsibility and business ethics "Yes" - Other concepts "Yes" Methodology integrates: - ESG criteria "Yes" - SEE (Social, Environmental & Ethical) risk rating "Yes" "The EIRIS methodology has drawn upon or reflects general initiatives such as the Global Reporting Initiative (GRI), OECD Guidelines for Multinational Enterprises and the UN Global Compact. For environmental areas the following initiatives have been incorporated: ICC Business Charter for Sustainable Development; CERES Principles; […]. For governance, key initiatives have included the UK Combined Code; OECD Principles of Corporate Governance, […]. Human and indigenous rights and supply chain issues have drawn upon the Universal Declaration on Human Rights; UN Human Rights Norms for Business; UN Global Compact; […]. Convention Watch has addressed the UN Human Rights Norms for Business; […]."
TC7	☐	☐	☐	–	

TC8	☑ ☑ ☐	E01-02 *Categories:* E18-28 "Environment, Social, Governance, ESG risk, Positive Products & Services, Convention Watch, Other Ethical Concerns" *E18 Community involvement; community impact:* "Community Involvement" *E19 Corporate Governance:* "Board Practice, Women on the Board, Board Level Accountability for Stakeholders, Bribery and Corruption, Countering Bribery, Code of Ethics, Convention Watch: Anti-Bribery Principles" *E20 Customer/product responsibility:* "Positive products and services, Relationships with Customers and Suppliers, Advertising complaints [UK only]" *E21 Environment:* "Environmental impact, Environmental policy, Environmental management, Environmental reporting, Environmental performance, Biodiversity, Chemicals of concern, Climate change, Greenhouse gases, Mining and quarrying, Nuclear power, Pollution convictions [UK only], Product Stewardship, Sustainable timber, Tropical forest clearance, Water pollution [UK only], Water use, Convention Watch: Kyoto Protocol, Convention Watch: Montreal Protocol, Convention Watch: Biodiversity, Convention Watch: Environmental Pollution (See also Environmental Solutions under Positive Screening Areas)." *E22 Health and safety:* "Health and Safety systems and prosecutions" *E23 Human Rights* "Human Rights - Country presence, Human Rights - large / small presence, Human Rights Overall, Human Rights Policy, Human Rights Systems, Human Rights Reporting, Indigenous rights; Convention Watch: International Human Rights Principles" *E24 Labour practices and decent work* "Equal Opportunities, Job Creation and Security, Trade Unions and Employee Participation, Training and Development" *E25 Social, Environmental, Ethical (SEE) Risks* "ESG Risk Management, Access to medicines, Chemical safety and sustainability, Mobile telecommunications health concerns, Obesity, Project finance and sustainability. Also risk / impact focus under environment, climate change, biodiversity, water use, bribery, human rights, and supply chain." *E26 Stakeholder Engagement* "Stakeholder Policy, Stakeholder Systems, Stakeholder Engagement, Stakeholder Reporting" *E27 Supply chain responsibility* "Global Sourcing exposure, Global Sourcing Overall, Global Sourcing Policy, Global Sourcing Systems, Global Sourcing Reporting, Convention Watch: International Labour Standards" *E28 Other* "Alcohol, Animal testing, Clinics – abortion, Cluster bombs, Developing World, Financial institutions, Fur, Gambling, Genetic engineering, Intensive farming and meat sale, Marketing breast milk substitutes, Military production and sale, International military sales activities, Military related turnover, Political do- nations [UK only], Pornography and adult entertainment services, Size, Tar sands & oil shale, Tobacco, Convention Watch: Ottawa Landmines treaty"
TC9	☑ ☐ ☐	E03	*E03 Criteria/Indicators:* "338"
TC10	☐ ☐ ☐	–	
TC11	☐[5] ☐[5] ☐[5]	–	

TC11a	☑	☑[5]	☐[5]	G05 G07-08 I05-06 "Assessments are produced in discreet areas e.g. assessments are made for each of policy, management systems, reporting for environment, climate change, equal opportunities, human rights, supply chain etc. Clients can then combine, weight, score, rank etc. these assessments in accordance with their own investment policies and priorities." "Clients can set up investment policies reflecting their concerns, areas of interests, weight of emphasis etc. in EGP [Eiris Global Platform] to allow them to adapt or combine EIRIS assessments, or convert them into numerical grades to either provide an alternative conclusion e.g. 'met' / 'not met' or achieve an overall score for the companies." "Analysts follow set procedures to produce assessments for most criteria"
TC11b	☐[5]	☐[5]	☐[5]	–
TC12	☑[5]	☐[5]	☐[5]	D40 I06 "Environmental Performance also gives different weight to different KPIs depending on the sectors of activity." "The flexibility of EGP gives clients a range of options. For instance they can give weightings to the EIRIS assessment and construct their own scores / rankings etc. in accordance with their own investment policies."
TC13	☑	☑	☐	D27 G15 "Over the last 13 yrs approx., EIRIS' methodology has changed significantly with the introduction of assessment grades in a number of areas, particularly around policies, management systems and reporting." "[Principles followed during assessment] are set out in the Guide to EIRIS Research and its companion Practical Guide to EIRIS Research. For many of the areas EIRIS researches, it adopts assessment grades, either qualitative (typically no evidence, limited, intermediate, good, advanced) or quantitative (e.g. 0-5%, 5-10%, 10-33%, >33% turnover). These provide an evaluation for each area / criteria. The clients can use these in any way they wish to produce their own assessment, for example, rankings, best in class, exclusions."
TC13a	☑	☐	☐	G15 "[Principles followed during assessment] are set out in the Guide to EIRIS Research and its companion Practical Guide to EIRIS Research."
TC14	☐[5]	☐[5]	☐[5]	–

Sources: (Eiris, 2012). Own presentation. R, D, and J refer to the three transparency levels reference, description, and justification.
Remarks: [1] decision is left to the customer. [5] only applicable to sub-phenomena, as final aggregation of indicators is left to the customer.

Ethifinance

Table A4.2: Elements disclosed by Ethifinance in its transparency matrix

#	R	D	J	Source	Referenced Extracts
TC1	☐	☐	☐	–	
TC2	☑[2]	☑	☐	–	
TC3	☑[1]	☐[1]	☐[1]	F26	"For our core methodology, the selection of companies is only based on clients request."
TC4	☑	☑	☐	F29-31 F63	Global operations are covered "Yes" Subsidiaries, joint ventures and associated companies are covered "Yes" Upstream ownership (shareholding) is analyzed "No" "We research company global operations on a consolidated scope, including joint ventures and subsidiaries. Our research methodology includes assessing the scope of information provided by each company on a specific aspect. Also we take into account the company country risk exposure as regards corruption and human rights issues meaning that before researching a company we identify those specific countries and then, based on a country risk exposure assessment, would have to respond to specific criteria."
TC5	☐	☐	☐	–	
TC6	☑[3]	☐	☐	D02-09 D39	Methodology is inspired by: - General international codes, conventions, guidelines "Yes" - International codes on corporate sustainability "Yes" - Concepts of corporate social responsibility and business ethics "Yes" - Other concepts "Yes" Methodology integrates: - ESG criteria "Yes" - SEE (Social, Environmental & Ethical) risk rating "Yes" "ILO Core Conventions, UDHR, several international conventions relating to environmental issues (Rotterdam, Montreal Protocol, Ramsar...), National regulations on corporate accountabil- ity (NRE Law in 2001 and LSF in 2003 in France), International guidelines so as OEDC Principles, ILO Tripartite Declaration, GRI guidelines + norms ISO 9001, 14001, 26000, SA 8000. As an example of Guidelines for Human Rights, we would base our research on Management Practices; […]. Looking at environmental issues, we would for instance take account of European Directives such as REACH, RoHS, WEEE and apply them to relevant business sectors being affected by such directives.
TC7	☐	☐	☐	–	
TC8	☑	☑	☐	E01-02 E18-28	*Categories:* "Governance, Social, Environmental and Other Stakeholders" *E18 Community involvement; community impact:* – *E19 Corporate Governance:* "Separation of Chairman and CEO, Board independence, Board practices, Board committees, Directors' remuneration policy, etc. Please refer to EthiFinance research framework for detail of indicators." *E20 Customer/product responsibility:* "Quality management and product responsibility issues policy and systems. Please refer to EthiFinance research framework for detail of indicators." *E21 Environment:* "Environmental policy, Environmental management, Energy consumption, Water consumption, Greenhouse gases, Water pollution, Waste Biodiversity, Chemicals, Soils pollution. Please refer to EthiFinance research framework for detail of indicators."

			E22 Health and safety: Health and safety risks identification and management, performance trend. Please refer to EthiFinance research framework for detail of indicators." *E23 Human Rights* "Human Rights Policy, monitoring and control, reporting. Please refer to EthiFinance research framework for detail of indicators." *E24 Labor practices and decent work* "Staff evolution (HR strategy, turnover, etc.), human capital (diversity, training, etc.), social dialogue. Please refer to EthiFinance research framework for detail of indicators." *E25 Social, Environmental, Ethical (SEE) Risks* "Responsibility, identification and management systems of non-financial risks. Please refer to EthiFinance research framework for detail of indicators." *E26 Stakeholder Engagement* "Identification of stakeholder, stakeholder mapping, taking into account stakeholders in the definition of the CSR strategy." *E27 Supply chain responsibility* "Sourcing policy, inclusion of social and environmental clauses in contracts, suppliers social audits, etc. Please refer to EthiFinance research framework for detail of indicators." *E28 Other* "Relations with shareholders (transparency of financial communication, operations on shares, ...). Please refer to EthiFinance research framework for detail of indicators."
TC9	☑	(☑)[4] ☐	E03 *E03 Criteria/Indicators:* "190" E18-28 See above for examples.
TC10	☐	☐ ☐	–
TC11	☐[5]	☐[5] ☐[5]	–
TC11a	☑	☑[5] ☐[5]	G07-08 "Yet the ratings are impacted by the weightings allocated by each client (on categories, aspects and criteria)." G15 "Final assessments are the results of a combination between the clients weightings and EthiFinance sectorial weightings. This is all explained in the EthiFinance Research Guide."
TC11b	☐[5]	☐[5] ☐[5]	–
TC12	☑[5]	☐[5] ☐[5]	D40 "[…] clients choose their own weightings of criteria assessed and have the ability to G13 ignore some criteria or some CSR themes (for example corporate governance)." G15 "Weightings are disclosed on company reports"
TC13	☑	☑ ☐	G08 "For every criteria, researchers select a standard answer based on the information G15 available on the researched company. […] Every standard answer relates to a grade from 0 to 1." "The output are quantitative (grades) and qualitative (written summaries) assessments. Then clients can use the results to make their own assessment (best-in-class, benchmarking, etc.)."
TC13a	☑	☐ ☐	G08 "Specific guidelines are disclosed for every criteria and the researcher has to justify the standard answer selected."
TC14	☐[5]	☐[5] ☐[5]	–

Sources: (EthiFinance, 2012). Own presentation. R, D, and J refer to the three transparency levels reference, description, and justification.
Remarks: [1] decision is left to the customer. [2] no exclusion criteria used. [3] list of international conventions. [4] partial examples only. [5] only applicable to sub-phenomena, as final aggregation of indicators is left to the customer.

Ges

Table A4.3: Elements disclosed by Ges in its transparency matrix

#	R	D	J	Source	Referenced Extracts
TC1	☐	☐	☐	–	
TC2	☑[1]	☐[1]	☐[1]	L31	Decisions on exclusions are left to customers: "Yes"
TC3	☑	☑	☐	F24-28 F64	"We monitor 18400 companies worldwide - MSCI All World, FTSE All World, Nordic indices and some client specific indices. Of these app. 4000 are rated on risk criteria - MSCI All World and Nordic indices."
TC4	☑	☑	☐	F29-31 F63	Global operations are covered "Yes" Subsidiaries, joint ventures and associated companies are covered "Yes" Upstream ownership (shareholding) is analyzed "Yes" "Upstream ownership is analysed in the process of norm violation assessment. If a company owns 20% or more of a company which has been found to vio- late international norms the owner company is included in the analysis and may be subject to an exclude/engage recommendation."
TC5	☐	☐	☐	–	
TC6	☑[3]	☐	☐	D02-09 D39	Methodology is inspired by: - General international codes, conventions, guidelines "Yes" - International codes on corporate sustainability "Yes" - Concepts of corporate social responsibility and business ethics "Yes" - Other concepts "Yes" Methodology integrates: - ESG criteria "Yes" - SEE (Social, Environmental & Ethical) risk rating "Yes" "UN Global Compact, OECD Guidelines for Multinational Enterprises, ILO Core Labour Conventions, Environmental Conventions, Human Rights Conventions and Weapons-related Conventions. Based on these we rate the companies and monitor for compliance."
TC7	☐	☐	☐	–	
TC8	☑	☑	☐	E01-02 E18-28	*Categories:* – *E18 Community involvement; community impact:* "I.e. Local community involvement" *E19 Corporate Governance:* "Board Management and control, Transparency & Incentive, Shareholder Rights" *E20 Customer/product responsibility:* "I.e. Policy and program" *E21 Environment:* "Environmental Management, Policy and Programs, Implementation of environmental management system" *E22 Health and safety:* "I.e. Policy and program" *E23 Human Rights* "I.e. Policy and program" *E24 Labour practices and decent work* "I.e. Policy and program" *E25 Social, Environmental, Ethical (SEE) Risks* "I.e. Policy and program" *E26 Stakeholder Engagement* "I.e. Policy and program" *E27 Supply chain responsibility* "I.e. Policy and program" *E28 Other* "None"

TC9	☑	(☑)[4]	☐	E03 D40	*E03 Criteria/Indicators:* "65" "A general criteria may be "Does the company have an environmental management system?" and an industry specific criteria may be "How much of raw material purchases have FSC Chain-of-Custody Certification?"
TC10	☐	☐	☐	–	
TC11	☐	☐	☐	–	
TC11a	☐	☐	☐	–	
TC11b	☐	☐	☐	–	
TC12	☐	☐	☐	–	
TC13	☐	☐	☐	–	
TC13a	☐	☐	☐	–	
TC14	☐	☐	☐	–	

Sources: (Ges, 2012). Own presentation. R, D, and J refer to the three transparency levels reference, description, and justification.
Remarks: [1] decision is left to the customer. [3] list of international conventions. [4] partial examples only.

Oekom

Table A4.4: Elements disclosed by Oekom in its transparency matrix

#	R	D	J	Source	Referenced Extracts
TC1	☐	☐	☐	–	
TC2	☑	☑	☐	L31 G15 E18	Decisions on exclusions are left to customers: "Yes" "In addition, the companies assessed are screened according to a wide range of exclusion criteria. oekom research's customers have the opportunity to either follow oekom research's pre-defined best-in-class approach, to adapt it to their individual perception of sustainability/CSR and to choose among the exclusion criteria those that should be applied to their universe." "Exclusion criteria: Violations of human rights; Controversial environmental practices"
TC3	☑	☑	☐	F24-27 F64	oekom research's Research Universe covers approximately 3,100 companies. It includes relevant conventional indexes (MSCI World, DJ Stoxx 600, MSCI Emerging Markets) as well as national indices (e.g. the DAX family, ATX, SMI, CAC40). Also included are sustainability leaders apart from those indexes, small and mid-caps from sectors with links to sustainability (e.g. renewable energies, recycling, water treatment), important non-listed bond issuers (mortgage bonds, corporate bonds, supranationals, governmental bonds) and titles included on specific customer requests. […] Of the 1060 companies listed above, about 810 are evaluated using the oekom Corporate Rating (Inside) methodology as such, which is described in detail in this transparency matrix. The remaining 250 companies are assessed via the oekom Corporate Rating (Outside) methodology […]."
TC4	☑	☑	☐	F29-31 F63	Global operations are covered "Yes" Subsidiaries, joint ventures and associated companies are covered "Yes" Upstream ownership (shareholding) is analyzed "Yes" "oekom research considers in its ratings the entire operations of the assessed companies worldwide. Regarding the inclusion of subsidiaries, joint ventures (JV) and associated companies in the assessment, oekom research defined the following standard approach: These are considered if the assessed company owns 50% or more of the voting rights or share capital; regarding any exclusion criteria, they are also included below this threshold if the investment can be seen as a "strategic" one, directly related to the business activity which is considered "controversial" of the subsidiary, JV or associated company. In this regard, JV are generally considered as a "strategic" investment. Based on a case to case decision process, subsidiaries may also be included below the 50% ownership limit."
TC5	☑	☐	☐	D39	"Originally, the rating methodology has been based on the Frankfurt-Hohenheim Guidelines (FHG), known as the most comprehensive set of criteria for the ethical/sustainable evaluation of companies. The FHG have been developed by a team of German scientists headed by Prof. Johannes Hoffmann. They include 800 indicators and are based on a value-tree analysis which resulted in three main dimensions: cultural sustainability, social sustainability and environmental sustainability. In cooperation with this project team, in 1999, oekom research transferred the guidelines into an implementable set of universal and sector-specific indicators. Using this rating structure, over 100 separate criteria have been selected […]"
TC6	☑[3]	☐	☐	D02-09 D39	Methodology is inspired by: - General international codes, conventions, guidelines "Yes" - International codes on corporate sustainability "Yes" - Concepts of corporate social responsibility and business ethics "Yes" - Other concepts "Yes" Methodology integrates: - ESG criteria "Yes" - SEE (Social, Environmental & Ethical) risk rating "Yes"

					"Frankfurt-Hohenheim Guidelines (FHG), known as the most comprehensive set of criteria for the ethical/sustainable evaluation of companies. The FHG have been developed by a team of German scientists headed by Prof. Johannes Hoffmann. They include 800 indicators and are based on a value-tree analysis which resulted in three main dimensions: cultural sustainability, social sustainability and environmental sustainability." "The FHG as well as oekom research's daily work are inspired by fundamental codes and principles regarding social/ethical and environmental standards. Important examples include the UN Declaration of Human Rights, the UN Global Compact, the core ILO conventions, as well as the OECD Guidelines for Multinational Enterprises. The influence of these standards can be tracked down to specific rating criteria, whose written methodology clearly outlines the consideration of these standards in the evaluation process. Also, they are reflected in specific exclusion criteria, which are designed along these standards, e.g. "Labour Rights" refers specifically to the four fundamental principles of the ILO Declaration on Fundamental Principles and Rights at Work, i.e. freedom of association, forced labour, child labour, discrimination."
TC7	☑	☐	☐	D39	"In co-operation with this [FHG] project team, in 1999, oekom research transferred the [Frankfurt Hohenheim] guidelines into an implementable set of universal and sector-specific indicators. Using this rating structure, over 100 separate criteria have been selected with the aid of which the complex circumstances of the respective areas under investigation can be adequately modelled. When the selection was made, special emphasis was placed on the functional context, relevance, data availability and data quality. In order to make the rating applicable to all countries worldwide, value was also attached to the universal relevance of criteria, irrespective of geographical location, development status, etc."
TC8	☑	☑	☐	E01-02 E18-28	*Categories:* "[…] These are structured in two dimensions, the Environmental and the Social Rating. The indicators assessed in the Environmental dimension of our oekom Corporate Rating (Inside) are grouped in three categories, i.e. Environ- mental Management, Products and Services, and Eco-efficiency; the three categories of the Social dimension are Staff and Suppliers, Society and Product Responsibility, and Corporate Governance and Business Ethics."
TC9	☑	☑	☐	E03 D40	*E03 Criteria/Indicators:* "119 - Estimate; based on one exemplary industry (Paper and Forest Products)" "Several of our indicators (including the criteria below) cover not only policy or management or performance or reporting, but two aspects (or more) at the same time. […] However, in order to respect the structure of the Transparency Matrix, each indicator was only counted once per line, and not two or more times as would have been appropriate" "The indicators below are exemplary and cover the Paper and Forest Products industry." *E18 Community involvement; community impact:* "- Community: […] Transparency and type of social, cultural and environmental corporate contributions during the previous year - Political donations - Transparency on participation in public policy making and lobbying activities - Taxes and subsidies: Transparency of payments to governments broken down by country; Transparency of financial assistance received from governments broken down by country - Stakeholder dialogue: External reporting on environmental issues; […] - Product responsibility: Measures taken to create opportunities for local communities and indigenous peoples to benefit from forestry operations"

E19 Corporate Governance:
"- Independence and effectiveness of the board: Separation of power between CEO and chairman of the board; […]
- Shareholder democracy: Voting rights and control enhancing mechanisms; Size of shareholding necessary to raise a resolution; Facilitation of shareholder participation
- Transparency of compensation schemes for executive board members
- Integration of sustainability performance aspects into the variable components of executive board members' remuneration
- Transparency of shareholder structure"
E20 Customer/product responsibility:
"-Product responsibility: Measures taken to create opportunities for local communities and indigenous peoples to benefit from forestry operations
- Product responsibility: Product safety management
- Product responsibility: Provision of customer/consumer information on environmental and social/ethical aspects of products - Product responsibility: Policy regarding responsible marketing
- Other major company-specific issues related to product and customer responsibility
- Major controversies, fines or settlements relating to customer and product responsibility
- Products and services - Forest/plantation management: Policy regarding sustainable forest/plantation management; Policy regarding genetically modified trees; Measures regarding sustainable forest/plantation management; Additional measures taken to actively protect/create habitat for diverse forest-based species; […]
- Products and services - Sourcing of fresh wood/fibre: Policy regarding sustainable sourcing of fresh wood/fibre; Measures taken to exclude procurement of fresh wood/fibre from unwanted sources; Percentage of (procured) fresh wood/fibre procured under a certified traceability system; […]
- Products and services - Sourcing of recycled wood/fibre: Strategy to increase the use of recycled wood/fibre; Percentage of recycled wood/fibre as of overall wood/fibre raw material
- Products and services - Wood/fibre processing: Measures taken to improve energy efficiency; Electricity produced through cogeneration; Energy use by source; […]
- Other major company-specific issues related to the integration of environmental considerations into products and services
- Major controversies, fines or settlements relating to environmental issues"
E21 Environment:
"The aspects "Product responsibility", "Products and services" and "Eco-efficiency" include a varying number of industry-specific indicators, covering the most relevant impacts of the industry's production processes, products, and services. The respective indicators below are exemplary and cover the Paper and Forest Products industry.
- Corporate policy covering environmental issues
- Environmental management system: Implementation of an environmental management system (EMS); Certification to an international standard
- External reporting on environmental issues
- Environmental performance indicators
- Strategy for addressing climate change and related risks: Policy; Responsibilities; Inventories; Targets and action plans; Transparency on risk exposure and adaptation/mitigation strategy: Physical risks, Regulatory risks; Market, cost and legal risks
- Travel and transport: Policy/measures taken to reduce the environmental impact of business travel; Measures taken to increase transport efficiency
- Suppliers: Subcontractor/supplier standards with regard to environmental issues; […]"
E22 Health and safety:
"Health and safety: Health and safety conditions in countries where the company operates; Health and safety management system; Accident / occupational illness rate; Number of fatal accidents during the last three years; Major controversies, fines or settlements relating to health and safety"

			E23 Human Rights "As indicators covering labour rights as well as the supply chain are counted separately, they have not been included here. oekom research's definition of human rights includes a company's relation to society and the community, but excludes a company's relation to its staff or suppliers. - Human rights: Policy on human rights of communities/peoples affected by the company's activities; Measures and methods to analyse and reduce potential negative social impact of company activities; Major controversies, fines or settlements relating to human rights […]" *E24 Labour practices and decent work* "- Freedom of association: Policy regarding freedom of association; Measures regarding freedom of association in countries where no independent labour union may be established or where freedom of association is prohibited by law; Major controversies, fines or settlements relating to freedom of association - Work-life balance: Options regarding reduction of working time and workplace flexibility; Options regarding dependent care and flexible benefits; Average weekly working time - Safeguarding of jobs: Occurrence of large-scale redundancies, significant job cuts or plant closures; Adjustment plans and measures to minimise the social impacts of large-scale redundancies, significant job cuts or plant closures - Payment: Minimum wages to ensure a decent standard of living for a worker and family; Staff benefits in low and middle income countries - Equal opportunities: Policy on equal opportunities; Gender distribution in the company; Major controversies, fines or settlements relating to discrimination - Training and education: Training and education policy and management; Transparency on average training hours per year per employee by employee category - Forced labour: Policy regarding forced labour; Major controversies, fines or settlements relating to forced labour - Child labour: Policy regarding child labour; Major controversies, fines or settlements relating to child labour" *E25 Social, Environmental, Ethical (SEE) Risks* "All of our indicators deal with a company's exposure, strategy and opportunities regarding social, environmental and ethical risks. To us, the SEE exposure of a company cannot be limited to single issues such as corporate governance or climate change. Rather, companies have to address a wide range of aspects and themes, in order to prepare truly for SEE challenges, current as well as possible future ones. Therefore, oekom research's Corporate Rating as a whole should be seen as covering SEE risks." *E26 Stakeholder Engagement* "- Suppliers: […] - Human rights: […] - Community: […] - External reporting on environmental issues *E27 Supply chain responsibility* "- Suppliers: Subcontractor/supplier standards with regard to labour / health and safety issues; Measures taken to check compliance of key subcontractors/suppliers with the company's labour / health and safety standards; Measures taken to support key suppliers in their labour / health and safety management; Major social controversies, fines or settlements relating to subcontractor/supplier activities - Suppliers: Subcontractor/supplier standards with regard to environmental issues; […] - Major controversies, fines or settlements relating to environmental issues" *E28 Other* "NR"
TC10	☐	☐ ☐	–
TC11	☑	☐ ☐	D40 "On the basis of a scientifically developed model, the weighting of the individual sections of the rating are adapted for each industry, depending on the specific risk profile of each industry."

TC11a	☑	☑	☐	G16 D38	"There is a Scientific Advisory Board which is consulted regarding major changes to the methodology, and which provides input on different aspects of the rating methodology. In addition, there is a Rating Committee, which deal with the concrete implications of our applied methodology. The external members of the Rating Committee are members of the Advisory Board and usually represent "sustainability practitioners", e.g. consultants or members of think tanks. […] Discussions may include the range of issues covered, the weighting of specific aspects for the industry or the operationalization of individual indicators."
TC11b	☐	☐	☐	–	
TC12	☑	☐	☐	G15	"The overall grade is generated by these individual scores according to their weighting."
TC13	☑	☑	☐	G15	"When evaluating information for individual scores, oekom research uses a grade between 1 (lowest grade) and 4 (highest grade). The overall grade is generated by these individual scores according to their weighting. Each numerical grade corresponds to a letter grade between D- (lowest grade) and A+ (highest grade), which is communicated. Within each industry, all oekom Corporate Ratings (Inside) of the companies assessed are compared and result in one overall ranking list of the industry. "
TC13a	☑	☐	☐	G15	"A handbook exists for each score, which defines the rating background as well as which performance/information relates to which grade. Additional guidance documents outline the overall research, assessment and evaluation framework and processes. […] oekom research follows its pre-defined best-in-class rating approach, which includes an industry-specific weighting of the different parts of the rating, an industry-specific minimum grade, as well as certain industry-specific criteria. This industry-specific framework is defined according to the industry's major challenges and exposure in terms of sustainability and based on a scientifically developed matrix."
TC14	☐	☐	☐	–	

Sources: (Oekom, 2012). Own presentation. R, D, and J refer to the three transparency levels reference, description, and justification.
Remarks: [3] list of international conventions.

Vigeo

Table A4.5: Elements disclosed by Vigeo in its transparency matrix

#	R	D	J	Source	Referenced Extracts
TC1	☐	☐	☐	–	
TC2	☑[1]	☐[1]	☐[1]	L31	Decisions on exclusions are left to customers: "Yes"
TC3	☑	☐	☐	F24-28 F62	"The pre-selection methodologies is based on the inclusion in both equity & bond indices. The capitalisation / the amount of debt on the bond market, the geographic localisations (Currently EUR, NAM and AP for the equity selection process), or the currency used when issuing debt are criteria that are taken into account. Two research universes are defined. In the first one, companies within the followed indices. In the other one, other companies."
TC4	☑	☑	☐	F29-31	Global operations are covered "Yes" Subsidiaries, joint ventures and associated companies are covered "Yes" Upstream ownership (shareholding) is analyzed "No"
TC5	☐	☐	☐	–	
TC6	☑[3]	☑	☐	D02-09 D39	Methodology is inspired by: - General international codes, conventions, guidelines "Yes" - International codes on corporate sustainability "Yes" - Concepts of corporate social responsibility and business ethics "Yes" - Other concepts "Yes" Methodology integrates: - ESG criteria "Yes" - SEE (Social, Environmental & Ethical) risk rating "Yes" "1. Human Resources: Continuous improvement of professional relations, labour relations and working conditions. Internationally recognised reference texts are amongst others: ILO: Tripartite Declaration of Principles concerning Multinational Enterprises and Social Policy ILO: […] 2. Human Rights at the Workplace: Respect for freedom of association, the right to collective bargaining, non-discrimination and promotion of equality, elimination of illegal working practices such as child or forced labour, prevention of inhumane or degrading treatment such as sexual harassment, protection of privacy and personal data. Internationally recognised reference texts are amongst others: UN: Universal Declaration of Human Rights ILO: ILO Declaration on Fundamental Principles and Rights at Work […] 3. Environment: Protection, safeguarding, prevention of damage to the environment, implementation of an adequate management strategy, eco-design, protection of biodiversity and coordinated management of environmental impacts on the entire lifecycle of products or services. Internationally recognised reference texts are amongst others: UN: Rio Declaration; EU: The Green Paper on Integrated Product Policy […] 4. Business Behaviour: Consideration of the rights and interests of clients, integration of social and environmental standards in the selection of suppliers and on the entire supply chain, effective prevention of corruption and respect for competitive practices. Internationally recognised reference texts are amongst others: UN: United Nations Guidelines for Consumer Protection, revised 1999; OECD: OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data […] 5. Corporate Governance: Effectiveness and integrity, guarantee of independence and efficiency of the Board of Directors, effectiveness and efficiency of auditing and control mecha- nisms, in particular the inclusion of social responsibility risks, respect for the rights of shareholders, particularly minority shareholders, transparency and rationale for the remu

					neration of directors. Internationally recognised reference texts are amongst others: OECD: OECD Corporate Governance Principles[…] 6. Community Involvement: Effectiveness, managerial commitment to community involvement, contribution to the economic and social development of territories / societies within which the company operates, positive commitment to manage the social impacts linked to products or services and overt contribution and participation in causes of public or general interest. Internationally recognised reference models are amongst others: WTO: Agreement on trade-related aspects of intellectual property rights (ADPIC) […]"
TC7	☐	☐	☐	–	
TC8	☑	☑	☐	E01-02	*Categories:* E18-28 "6 - Human Resources, Human Rights, Community Involvement, Environment, Business Behaviour, Corporate Governance" *E18 Community involvement; community impact:* "Promotion of social and economic development, Societal impacts of products/ service, Contribution to general interest causes" *E19 Corporate Governance:* "Board of Directors, Audit& Internal Controls, Shareholders, Executive Remuneration" *E20 Customer/product responsibility:* "Product safety, Information to customers, Responsible customer relations" *E21 Environment:* "Environmental strategy and eco-design, Pollution prevention and control (soil, accident), Development of green products and services, Protection of biodiversity, Protection of water resources, Minimizing environmental impacts from energy use, Management of atmospheric emissions, Waste management, Management of local pollution, Management of environmental impacts from transportation, Management of environmental impacts from the use and disposal of products/services" *E22 Health and safety:* "Improvement of health and safety conditions including Stress at work" *E23 Human Rights* "Respect for human rights standards and preventions of violations, Respect for freedom of association and the right to collective bargaining, Non-Discrimination, Elimination of forced labour and child labour" *E24 Labour practices and decent work* "Promotion of labour relations, Encouraging employee participation, Responsible management of restructurings, Career management and promotion of employability, Quality of remuneration systems, Respect and management of working hours" *E25 Social, Environmental, Ethical (SEE) Risks* "Vigeo defines 4 risk classes: Reputation, Operational efficiency, Human capital, Legal security" *E26 Stakeholder Engagement* "There are no separate criteria on stakeholder engagement, but the level of stakeholder engagement is integrated in the rating scale of different criteria." *E27 Supply chain responsibility* "Sustainable relations with suppliers, Integration of environmental factors in the supply chain, Integration of social factors in the supply chain" *E28 Other* "Business integrity: Prevention of corruption, Prevention of anti-competitive practices, Transparency and integrity of influence strategies & practices."
TC9	☑	☐	☐	E03	*E03 Criteria/Indicators:* "305"
TC10	☐	☐	☐	–	
TC11	☐[5]	☐[5]	☐[5]	–	
TC11a	☐[5]	☐[5]	☐[5]	–	
TC11b	☐[5]	☐[5]	☐[5]	–	

TC12	☑[5]	☐[5]	☐[5]	D40	"The same standard set of procedures, criteria and rating guidelines applies to all companies that are analysed under the Equitics methodology, but for each sector, the weight and the precise content of each of the criteria is adjusted and customised."
TC13	☑	☑	☐	G04 G15	"The Equitics methodology does not include final assessments in the sense of judgments of which company is (not) 'good enough' for - for instance inclusion in an SRI fund. However, (non-final) assessments are made in the sense that every company gets for each of the six domains of CSR, scores and a rating (from double minus to double plus) that compares the performance of the company in that specific domain with the performance of its sector peers. These (non-final) assessment principles and procedures are described in this document." "Outline of principles followed: Scoring is conducted at criterion level (based on a predefined rating scale) and scores are integrated at domain level. The scorings range from 0 to 100. Ratings are conducted at domain level. The ratings rank from -- (double minus) to ++ (double plus) and indicate the position of a company on a given domain as compared to the sector."
TC13a	☑	☐	☐	G12	"The scores that are attached to specific performances or characteristic of a company are indicated by the 'scoring sentences' in the database."
TC14	☐[5]	☐[5]	☐[5]	–	

Sources: (Vigeo, 2012). Own presentation. R, D, and J refer to the three transparency levels reference, description, and justification.
Remarks: [1] decision is left to the customer. [3] list of international conventions. [5] only applicable to sub-phenomena, as final aggregation of indicators is left to the customer.

The end.