
Zusammenhang zwischen Nachhaltigkeitsperformance und Nachhaltigkeitsberichterstattung – Legitimität oder finanzielle Überlegungen?



Katrin Hummel und Christian Schlick

Nachhaltigkeitsberichterstattung, Nachhaltigkeitsperformance, Legitimacy Theory, Voluntary Disclosure Theory

Sustainability disclosure, sustainability performance, legitimacy theory, voluntary disclosure theory



Der Zusammenhang zwischen Nachhaltigkeitsberichterstattung und Nachhaltigkeitsperformance ist zum aktuellen Zeitpunkt weder theoretisch noch empirisch eindeutig erklärt. Theoretische Erklärungsansätze liefern sowohl die Legitimacy Theory als auch die Voluntary Disclosure Theory. Angesichts methodischer Schwächen bisheriger Studien bei der Operationalisierung der zentralen Untersuchungsvariablen steht im Mittelpunkt des Beitrags die Entwicklung neuer Konzepte zur Messung von Umfang und Form der Nachhaltigkeitsberichterstattung sowie der Nachhaltigkeitsperformance. Die entwickelten Messkonzepte eignen sich zur konsistenten Überprüfung der theoretischen Erklärungsansätze. Anhand einer Stichprobe von 50 Deutschen und Schweizer Grossunternehmen kann im Unterschied zu bisherigen Untersuchungen gezeigt werden, dass beide Theorien jeweils einen Teil der Nachhaltigkeitsberichterstattung erklären können.

Both theoretical considerations as well as empirical findings report mixed results on the relationship between sustainability disclosure and sustainability performance. Theoretical considerations are mainly based on legitimacy theory and voluntary disclosure theory. Due to shortcomings of previous studies with regard to the measurement of the main variables, our paper concentrates on new measurement approaches for sustainability disclosure and sustainability performance. Findings from a sample of 50 German and Swiss companies support our methodological approach. In particular, we find empirical evidence that legitimacy theory and voluntary disclosure theory likewise can explain certain aspects of sustainability reporting.

1. Problemstellung

Die Frage, warum Unternehmen freiwillig Informationen zur Nachhaltigkeitsperformance im Rahmen eines Geschäftsberichts oder eines separaten Nachhaltigkeitsberichts veröffentlichen, ist zum aktuellen Zeitpunkt weder theoretisch noch empirisch eindeutig beant-

wortet. Von besonderem Interesse ist dabei der Zusammenhang zwischen Nachhaltigkeitsperformance und Nachhaltigkeitsberichterstattung, wobei zwei (vermeintlich) divergierende theoretische Erklärungsansätze diskutiert werden. Der *Legitimacy Theory* (LT) folgend wird argumentiert, dass die Nachhaltigkeitsberichterstattung eingesetzt wird, um die öffentliche Wahrnehmung der (tatsächlich schlechten) Nachhaltigkeitsperformance eines Unternehmens positiv zu beeinflussen (vgl. exemplarisch *Deegan* 2002). Daher wird in empirischen Untersuchungen ein *negativer* Zusammenhang zwischen der Nachhaltigkeitsperformance und dem Umfang der Nachhaltigkeitsberichterstattung oftmals mit Argumenten der LT erklärt (vgl. exemplarisch *Dawkins/Fraas* 2011, 315f. und *Patten* 2002, 772). Demgegenüber argumentieren Vertreter der *Voluntary Disclosure Theory* (VDT) mit preissteigernden Effekten positiver Nachrichten am Kapitalmarkt und postulieren einen *positiven* Zusammenhang zwischen der Nachhaltigkeitsperformance und dem Umfang bzw. der Qualität der Nachhaltigkeitsberichterstattung (vgl. exemplarisch *Clarkson et al.* 2008).

Die Ergebnisse bisheriger empirischer Untersuchungen mehrheitlich US-amerikanischer Unternehmen fallen konträr aus und sind insofern wenig geeignet, die Fragestellung eindeutig zu beantworten. Sie konzentrieren sich ausschließlich auf den Umweltbereich als einen Teilbereich der Nachhaltigkeit. Insbesondere weisen sämtliche Untersuchungen bei der Messung und Operationalisierung der zentralen Variablen Schwächen auf, welche die Interpretation der Ergebnisse stark einschränken. So erfolgt nur bei wenigen Untersuchungen eine Unterscheidung zwischen quantitativer und verbaler Form der Berichterstattung. Diese Trennung ist jedoch zur Analyse der Problemstellung notwendig. Denn der Logik der LT folgend wird bei schlechter Nachhaltigkeitsperformance zwar umfangreicher berichtet, die Darstellung erfolgt aber bei kritischen Themen eher verbal. Hingegen basiert die VDT auf der Entscheidungsrelevanz der publizierten Informationen, was ebenfalls eine Unterscheidung von verbaler und quantitativer Information nahelegt. Des Weiteren fällt eine grosse Heterogenität bei der Messung und Operationalisierung der Nachhaltigkeitsperformance auf. Zur Operationalisierung wird mehrheitlich auf lediglich ein oder zwei isolierte Kenngrößen, wie beispielsweise Emissionsdaten oder Daten zur Abfallmenge, zurückgegriffen (vgl. exemplarisch *Clarkson et al.* 2008). Eine solche Messung der Umweltperformance ist hinsichtlich ihrer Validität fragwürdig und daher für eine grundlegende Beurteilung der Umweltpflege wenig geeignet.

Vor diesem Hintergrund besteht das Ziel dieses Beitrags in der Entwicklung und Anwendung von Konzepten zur Messung von Umfang und Form der Nachhaltigkeitsberichterstattung sowie der Nachhaltigkeitsperformance. Im Unterschied zu bisherigen Untersuchungen wird bei der Messung der Nachhaltigkeitsberichterstattung streng zwischen quantitativer und verbaler Form der Berichterstattung differenziert. Es wird ein valides Konzept zur Messung der Nachhaltigkeitsperformance vorgestellt. Dieses bietet wertvolle Ansatzpunkte zur Überwindung der Schwächen bisheriger Ansätze. Die Anwendung dieser Messkonzepte erlaubt belastbare Rückschlüsse auf die Gültigkeit der theoretischen Erklärungsansätze. Für ein Sample von 50 Deutschen und Schweizer Grossunternehmen wird gezeigt, dass beide Theorien jeweils einen *Teil* der Nachhaltigkeitsberichterstattung erklären können.

Der Beitrag gliedert sich in fünf Kapitel. Aufbauend auf einem umfassenden Überblick über bisherige empirische Untersuchungen des Zusammenhangs von Nachhaltigkeitsberichterstattung und -performance und einer Darstellung der theoretischen Erklärungsansätzen

ze werden im zweiten Kapitel die Forschungshypothesen der Untersuchung formuliert. Diese bilden den Ausgangspunkt für die Entwicklung von Messkonzepten zur Bestimmung von Umfang und Form der Nachhaltigkeitsberichterstattung sowie der Nachhaltigkeitsperformance, die im dritten Kapitel vorgestellt und im Hinblick auf die Erfüllung zentraler Gütekriterien kritisch reflektiert werden. Das vierte Kapitel beinhaltet die empirische Anwendung der Messkonzepte sowie die Überprüfung der formulierten Forschungshypothesen. Der Beitrag schliesst mit einem Fazit und einer kritischen Diskussion.

2. Aktueller Forschungsstand, theoretische Fundierung und Hypothesenformulierung

Tabelle 1 gibt einen Überblick über bisherige empirische Untersuchungen des Zusammenhangs von Nachhaltigkeitsberichterstattung und -performance. Gemäss des Triple-Bottom-Line-Ansatzes umfasst Nachhaltigkeit die drei Säulen Ökonomie, Ökologie und Soziales (vgl. Elkington 1999). Im Kern geht es um die ökonomische, ökologische und soziale Verantwortung, welche eine Unternehmung gegenüber ihrem Umfeld wahrzunehmen hat.

Autor (Jahr)	n Land	Messung Berichterstattung	Messung Performance	Ergebnis
Cho et al. (2012)	92 US	<ul style="list-style-type: none"> ■ Analyse Nachhaltigkeits- und Geschäftsberichte ■ vgl. Clarkson et al. (2008) 	<ul style="list-style-type: none"> ■ Environmental Impact Score von Newsweek: u.a. Luftemissionen, Wassernutzung, Abfälle 	-
Clarkson et al. (2011)	51 AU	<ul style="list-style-type: none"> ■ Analyse Nachhaltigkeits- und Geschäftsberichte ■ vgl. Clarkson et al. (2008) 	<ul style="list-style-type: none"> ■ giftige Freisetzung pro Umsatz (90 risikogewichtete Substanzen) 	-
Dawkins/ Fraas (2011)	344 US	<ul style="list-style-type: none"> ■ Reaktionen auf Umfrage des Carbon Disclosure Projects ■ 3 Gruppen: keine Antwort/Ablehnung, unvollständige Antwort/Verweise, vollständige Beantwortung 	<ul style="list-style-type: none"> ■ KLD-Rating: Stärken und Schwächen ■ Rating der CO₂-Emissionen 	-
Clarkson et al. (2008)	191 US	<ul style="list-style-type: none"> ■ Analyse Nachhaltigkeitsberichte ■ 7 Kategorien mit 95 Elementen in Anlehnung an GRI (16 soft/leicht imitierbar und 79 hard/schwer imitierbar) ■ 0/1-Punkteschema 	<ul style="list-style-type: none"> ■ Recyclingrate giftiger Abfälle (Rang) ■ Giftige Abfälle pro Umsatz (Rang) 	+
Cho/Patten (2007)	100 US	<ul style="list-style-type: none"> ■ Analyse Geschäftsberichte ■ vgl. Patten (2002) ■ zusätzlich: getrennte Auswertung der 4 finanziellen und 4 nicht-finanziellen Elemente 	<ul style="list-style-type: none"> ■ KLD-Rating: Einteilung in schlechter/besser, sofern Schwächen im Umweltbereich identifiziert/nicht identifiziert werden 	-
Cho et al. (2006)	119 US	<ul style="list-style-type: none"> ■ Analyse Geschäftsberichte ■ vgl. Patten (2002) 	<ul style="list-style-type: none"> ■ KLD-Rating: Punktsumme über umweltbezogene Schwächen 	-
de Villiers/ van Staden (2006)	≥137 ZA	<ul style="list-style-type: none"> ■ Analyse Geschäftsberichte ■ in Anlehnung an Wiseman (1982) ■ Trennung allgemeine/spezifische Angaben ■ 0/1-Punkteschema ohne Gewichtung ■ getrennte Betrachtung von 8 allgemeinen Elementen der Mitarbeiterbezogenen Berichterstattung 	<ul style="list-style-type: none"> ■ keine explizite Messung ■ Zusammenhang wird unterstellt 	-

Autor (Jahr)	n Land	Messung Berichterstattung	Messung Performance	Ergebnis
Al-Tuwaijri et al. (2004)	198 US	<ul style="list-style-type: none"> ■ Analyse Geschäftsberichte ■ 4 Elemente: giftiger Abfall, Geldstrafen, Beseitigung giftiger Altlasten, Öl- und Chemieunfälle ■ Punkteschema: quantitativ (3), verbal spezifisch (2), verbal unspezifisch (1), keine Berichterstattung (0) ■ Relevanz für individuelles Unternehmen wird berücksichtigt 	<ul style="list-style-type: none"> ■ Recyclingrate giftiger Abfälle ■ Giftiger Abfall pro Umsatz ■ Haltung zu Umweltbelangen (3 Faktoren) 	+
Patten (2002)	131 US	<ul style="list-style-type: none"> ■ Analyse Geschäftsberichte ■ 4 verbale (Diskussion) und 4 quantitative (Kosten) Elemente ■ 0/1-Punkteschema ■ zusätzlich: Anzahl publizierter Zeilen 	<ul style="list-style-type: none"> ■ Menge freigesetzter gefährlicher Stoffe pro Umsatz 	-
Hughes et al. (2001)	51 US	<ul style="list-style-type: none"> ■ Analyse Geschäftsberichte ■ vgl. Wiseman (1982) ■ zusätzlich 5 neue Elemente ■ Punkteschema: quantitativ (4), verbal (3), oberflächlich (2), unerheblich (1), keine (0) 	<ul style="list-style-type: none"> ■ 3 Gruppen (gut, gemischt, schlecht) auf Basis eines Rankings des CEP (Council on Economic Priorities): Emissionen in Luft und Wasser, Vermeidungsprogramme 	-
Bewley/Li (2000)	188 CA	<ul style="list-style-type: none"> ■ Analyse Geschäftsberichte ■ vgl. Wiseman (1982) ■ zusätzlich: Trennung finanzielle/nicht-finanzielle Angaben 	<ul style="list-style-type: none"> ■ Verschmutzungsneigung (hoch/niedrig je nach Branche) ■ Meldung giftiger Freisetzung (ja/nein) 	+
Fekrat et al. (1996)	26 US	<ul style="list-style-type: none"> ■ Analyse Geschäftsberichte ■ vgl. Wiseman (1982) 	<ul style="list-style-type: none"> ■ vgl. Ingram/Frazier (1980) 	?
Rockness (1985)	26 US	<ul style="list-style-type: none"> ■ indirekte Messung als Wahrnehmung der Angaben in Geschäftsberichten durch Probanden 	<ul style="list-style-type: none"> ■ vgl. Ingram/Frazier (1980) 	?
Wiseman (1982)	26 US	<ul style="list-style-type: none"> ■ Analyse Geschäftsberichte ■ 4 Kategorien mit 18 Elementen: ökonomisch, Rechtsstreitigkeiten, Verschmutzung, andere Angaben ■ Punktbewertungsschema: quantitativ (3), verbal (2), allgemein (1), keine Berichterstattung (0) ■ zusätzlich: Anzahl publizierter Zeilen 	<ul style="list-style-type: none"> ■ vgl. Ingram/Frazier (1980) 	?
Ingram/ Frazier (1980)	40 US	<ul style="list-style-type: none"> ■ Analyse Geschäftsberichte ■ 4 Kategorien mit insgesamt 20 Ausprägungen: Art (u.a. finanziell, nicht-finanziell, qualitativ), Zeitbezug (Vergangenheit, Gegenwart, Zukunft), Genauigkeit (spezifisch, allgemein), Thema ■ Einzelprüfung Satz für Satz ■ Auswertung je Ausprägung und je Kategorie (Summe) 	<ul style="list-style-type: none"> ■ Index des CEP (Council on Economic Priorities): Emissionen in Luft und Wasser, Vermeidungsprogramme 	?

AU Australien CA Kanada US USA ZA Südafrika

+ positiver Zusammenhang - negativer Zusammenhang ? unklarer Zusammenhang

Tabelle 1: Überblick über bisherige empirische Untersuchungen des Zusammenhangs von Nachhaltigkeitsberichterstattung und Nachhaltigkeitsperformance

Sämtliche Untersuchungen in *Tabelle 1* konzentrieren sich auf den Umweltbereich und nehmen keine umfassende Analyse der Nachhaltigkeitsperformance vor. Zwar gibt es auch empirische Untersuchungen der Nachhaltigkeitsberichterstattung, allerdings bleibt bei diesen Untersuchungen die Nachhaltigkeitsperformance unberücksichtigt (vgl. exemplarisch Quick/Knocinski 2006; van der Laan Smith et al. 2005). Für die Ableitung möglicher Zusammenhänge zwischen Nachhaltigkeitsberichterstattung und Nachhaltigkeitsperformance werden in der Regel die Erklärungsansätze der VDT oder der LT herangezogen.

Gemäss der VDT, welche sich ursprünglich auf die Finanzberichterstattung am Kapitalmarkt bezieht, wird das freiwillige Publikationsverhalten von Unternehmen durch die Information selbst und die anfallenden Kosten der Publizität bestimmt (vgl. Verrecchia 1983). Die publizierte Information schlägt sich am Markt direkt im Unternehmenswert nieder. Es kann formal eine kritische Schwelle bestimmt werden, bei der die anfallenden Kosten der Publizität exakt der erwarteten Änderung des Unternehmenswertes aufgrund der Informationsbekanntgabe entsprechen. Informationen oberhalb dieser kritischen Schwelle werden berichtet, Informationen unterhalb dieser kritischen Schwelle werden nicht ausgewiesen. Zusammengefasst bietet das Modell eine Erklärung dafür, warum es aus Sicht eines Unternehmens rational sein kann, Informationen ab einem bestimmten Grenzwert ihrer Ausprägung nicht zu berichten. Diese Argumentationslogik wurde von verschiedenen Vertretern auf die freiwillige Berichterstattung nicht-finanzieller Informationen erweitert (vgl. exemplarisch Clarkson et al. 2008; Bewley/Li 2000; Li et al. 1997). Es wird argumentiert, dass nicht-finanzielle Informationen immer dann berichtet werden, wenn deren preisseigernde Effekte grösser sind als die damit verbundenen Kosten der Berichterstattung. Im Mittelpunkt steht somit die Entscheidung der Berichterstattung an sich, nicht jedoch in welcher Form berichtet wird. Der Argumentation der VDT folgend wird daher *Hypothese H₁* formuliert:

H₁: Es besteht ein positiver Zusammenhang zwischen der Nachhaltigkeitsperformance und dem Umfang berichteter Nachhaltigkeitsinformationen.

Implizit wird bei der VDT Entscheidungsrelevanz der berichteten Informationen unterstellt. In Anlehnung an Studien jüngeren Datums (vgl. Clarkson et al. 2008; Cho/Patten 2007; Al-Tuwairjri et al. 2004; Bewley/Li 2000) wird daher in einem zweiten Schritt die berichtete Nachhaltigkeitsinformation dahingehend differenziert, ob es sich um eine verbale oder quantitative Angabe handelt. Es wird argumentiert, dass insbesondere quantitative Informationen entscheidungsrelevant sind, da diese im Vergleich zu einer rein verbalen Berichterstattung eine bessere Objektivität und Vergleichbarkeit sowie eine geringere Manipulationsanfälligkeit aufweisen (vgl. Al-Tuwairjri et al. 2004, 454). Um adverse Selektion zu vermeiden, signalisieren Unternehmen mit guter Nachhaltigkeitsperformance ihren Zustand, indem sie quantitative Nachhaltigkeitsinformationen publizieren (vgl. Clarkson et al. 2008, 307f.). So können sie sich von Unternehmen mit schlechter Nachhaltigkeitsperformance unterscheiden. Zudem kann die Publikation einer quantitativen Information von Unternehmen mit schlechter Nachhaltigkeitsperformance nicht ohne Weiteres imitiert werden (vgl. Clarkson et al. 2008, 308). Diesem Erklärungsansatz folgend wird daher vermutet, dass Unternehmen mit besserer Nachhaltigkeitsperformance Informationen vermehrt quantitativ berichten. Diese Argumentation schlägt sich in folgender Hypothese nieder:

H_{2a}: Es besteht ein positiver Zusammenhang zwischen der Nachhaltigkeitsperformance und dem Ausmass der quantitativen Nachhaltigkeitsberichterstattung.

Einen anderen Erklärungsansatz bietet die LT, wonach der Fortbestand eines Unternehmens von der Legitimierung durch sein Umfeld abhängig ist (vgl. *Davis* 1973, 314). Dabei beschreibt Legitimität einen Zustand, bei welchem sich die Wertevorstellungen von Unternehmen und gesellschaftlichem Umfeld decken (vgl. *Dowling/Pfeffer* 1975, 122). Umgekehrt verwendet *Sethi* (1978, 58) den Begriff der Legitimitätslücke, sofern diese voneinander abweichen und bestimmte Unternehmensaktivitäten von der Gesellschaft als Normverletzung angesehen werden. Eine Wahrnehmung als Normverletzung ist jedoch auch dann möglich, wenn die Wertvorstellungen von Unternehmen und Gesellschaft übereinstimmen. Durch Informationsasymmetrien zwischen Unternehmen und Gesellschaft könnte die Gesellschaft eine tatsächlich wertkonforme Unternehmensaktivität fälschlicherweise als Normverletzung ansehen. In diesem Zustand der Legitimitätslücke kann dem Unternehmen die Versorgung mit personellen und materiellen Ressourcen durch die Gesellschaft erschwert oder verweigert werden. Um eine Legitimitätslücke zu schliessen, soll das Unternehmen durch die gezielte Bereitstellung und vorteilhafte Darstellung neuer Informationen die vom Umfeld wahrgenommene Unternehmensleistung zu Gunsten des Unternehmens beeinflussen (vgl. *Milne/Patten* 2002, 381; *Gray et al.* 1995, 54; *Sethi* 1978, 58). Nur wenn diese Strategie nicht erfolgreich ist, wird die Anpassung der tatsächlichen Unternehmensaktivität empfohlen (vgl. *Sethi* 1978, 58). Neben der Wiederherstellung (repair) von Legitimität existieren weitere Überlegungen zu ihrem Aufbau (gain) und Erhalt (maintain) (vgl. *O'Donovan* 2002, 349; *Suchman* 1995, 585ff.), wobei einem Unternehmen der präventive Erhalt der Legitimität nahegelegt wird (vgl. *Ashforth/Gibbs* 1990, 183). Übertragen auf die Nachhaltigkeitsberichterstattung bedeutet dies, dass Unternehmen Informationen zur Nachhaltigkeit ihrer Tätigkeit als ein Mittel zur gezielten Beeinflussung ihrer Legitimität nutzen können (vgl. *Deegan* 2002). Es wird vermutet, dass eine solch gezielte Beeinflussung der von der Gesellschaft wahrgenommenen Nachhaltigkeitsperformance insbesondere durch Publikation *verbaler* Informationen erfolgt, da nur bei diesen eine gezielt vorteilhafte Darstellung überhaupt möglich ist. Sowohl die Vergleichbarkeit als auch die Objektivität dieser Form der Information sind stark eingeschränkt. Je schlechter die Nachhaltigkeitsperformance eines Unternehmens ausfällt, desto eher wird es die klar nachvollziehbare quantitative Berichterstattung bei kritischen Sachverhalten durch eine verbale Darstellung ersetzen, um eine existierende oder drohende Legitimitätslücke zu schliessen. Dies führt zu folgender Hypothese:

H_{2b}: Es besteht ein negativer Zusammenhang zwischen der Nachhaltigkeitsperformance und dem Ausmass der verbalen Nachhaltigkeitsberichterstattung.

Eine Analyse der Studien aus *Tabelle 1* zeigt, dass in bisherigen Studien die Messung und Operationalisierung der dort interessierenden Variablen Umweltberichterstattung und Umweltperformance sehr unterschiedlich vorgenommen werden.

Für die Messung der *Umweltberichterstattung* hat sich die Inhaltsanalyse als methodischer Standard herausgebildet, wenngleich die berücksichtigten Indikatoren und die Art der Erfassung stark variieren. Für die Auswertung wird in der Regel auf eine aggregierte Gesamtbewertung zurückgegriffen. Lediglich bei einigen Studien jüngeren Datums werden ansatzweise quantitative und qualitative Information unterschieden (vgl. *Clarkson et al.* 2008; *Cho/Patten* 2007; *Al-Tuwarijri et al.* 2004; *Bewley/Li* 2000). Allerdings ist diese

Unterscheidung weder bei der Messung noch bei der Auswertung durchgängig umgesetzt.¹ Dies könnte darin begründet sein, dass bisherige empirische Untersuchungen darauf abziehen, die Gültigkeit entweder der einen *oder* der anderen theoretischen Argumentation zu festigen, indem der Zusammenhang zwischen Nachhaltigkeitsberichterstattung und Nachhaltigkeitsperformance untersucht wird. Ein positiver Zusammenhang wird als Beleg für die Gültigkeit der VDT, ein negativer Zusammenhang als Beleg für die Gültigkeit der LT interpretiert. Bei dieser Herangehensweise bleibt unbeachtet, dass beide Theorien für die Erklärung unterschiedlicher Aspekte der Nachhaltigkeitsberichterstattung gleichermaßen Bestand haben können. So schliessen Clarkson *et al.* (2008, 325) aus den Ergebnissen ihrer empirischen Untersuchung, dass die LT zusätzlich zur VDT geeignet zu sein scheint, die Umweltberichterstattung in bestimmten Situationen zu erklären.

Deutlich heterogener fällt die Messung der *Umweltpfifomance* aus. Häufig wird zur Operationalisierung auf lediglich ein oder zwei isolierte Kenngrößen, wie beispielsweise Emissionsdaten oder Daten zur Abfallmenge, zurückgegriffen. Ein solcher Ansatz erscheint im Hinblick auf die Validität der Messung fragwürdig. Alternativ wird in einigen Studien zur Messung der Umweltpfifomance auf die Beurteilung durch externe Ratingagenturen zurückgegriffen, wobei das Rating von *Kinder, Lydenberg, Domini* (KLD) am weitesten verbreitet ist (vgl. Dawkins/Fraas 2011; Cho/Patten 2007; Cho *et al.* 2006). Beim KLD-Rating wird die Umweltpfifomance eines Unternehmens anhand von 14 Indikatoren beurteilt. Pro Unternehmen werden sieben umweltbezogene Stärken und umweltbezogene Schwächen analysiert und binär (zutreffend/nicht zutreffend) codiert (vgl. Chatterji *et al.* 2009, 132ff.). Die Aggregation dieser Indikatoren ergibt ein Mass für die Umweltpfifomance eines Unternehmens. Auch wenn das KLD-Rating die Problematik der isolierten Berücksichtigung einzelner Aspekte der Umweltpfifomance überwindet, offenbart es gleichwohl Schwächen. Diese liegen zum einen in der Intransparenz, da konkrete, quantitative Messwerte sowie die Grenzwerte der binären Bewertung nicht nachvollziehbar sind. Zum anderen stellt die binäre Erfassung an sich eine starke Vereinfachung dar, da Streuungen innerhalb der Gruppen nicht erfasst werden. Außerdem liegen die Ratings von KLD überwiegend für US-amerikanische Unternehmen und weniger für europäische Unternehmen vor.

Die knappe Diskussion der methodischen Ansätze bisheriger Untersuchungen zeigt, dass sich in der Forschung bislang keine Standards für eine differenzierte Messung von Nachhaltigkeitsberichterstattung und Nachhaltigkeitsperformance herausgebildet haben. Bisherige Ansätze konzentrieren sich überwiegend auf den Umweltbereich und weisen zudem Schwächen auf, weshalb sie für die Hypothesenprüfung dieser Untersuchung ungeeignet sind. Im folgenden Abschnitt werden daher erweiterte Messkonzepte für die Nachhaltigkeitsberichterstattung und die Nachhaltigkeitsperformance entwickelt.

3. Untersuchungsmethodik

3.1 Gütekriterien der Messung

Die Messung von Nachhaltigkeitsberichterstattung und Nachhaltigkeitsperformance sollte sich an den in der einschlägigen Literatur angeführten Gütekriterien Objektivität, Reliabi-

1 Eine Ausnahme hiervon stellt die Untersuchung von Cho und Patten (2007) dar, in welcher der Zusammenhang zwischen Umweltpfifomance und Umweltberichterstattung für monetäre und nicht-monetäre Berichterstattung getrennt untersucht wird.

lität und Validität eines Messinstruments orientieren (vgl. *Bortz/Döring* 2006, 195ff.). Darüber hinaus erfordert der spezifische Untersuchungskontext eine Kongruenz der den beiden Messkonzepten zugrunde liegenden Informationsitems sowie eine mehrdimensionale Messung. Das Kriterium der *Objektivität* verlangt nach der Unabhängigkeit der Messung der interessierenden Variablen vom Beobachter (vgl. für die folgenden Ausführungen *Bortz/Döring* 2006, 195ff.). Mithin sollen verschiedene Anwender zu den gleichen Resultaten gelangen. Die *Reliabilität* ist ein Mass für die Zuverlässigkeit und Stabilität eines Messinstruments. Konkret sollen wiederholte Messungen dieselben Werte liefern. Die *Validität* beschreibt die Gültigkeit eines Messinstruments und untersucht, ob ein Messinstrument tatsächlich das misst, was gemessen werden soll. Die Anforderung der *inhaltlichen Kongruenz* bezieht sich auf die Gegenüberstellung von Nachhaltigkeitsberichterstattung und Nachhaltigkeitsperformance. Insbesondere im Hinblick auf die Überprüfung der VDT ist nach Möglichkeit eine Deckungsgleichheit der erhobenen Informationsitems der beiden Messkonzepte erforderlich. Das Kriterium der *Mehrdimensionalität* folgt aus dem Triple-Bottom-Line-Ansatz und verlangt, dass möglichst alle Dimensionen der Nachhaltigkeit sowie verschiedene Aspekte der einzelnen Dimensionen der Nachhaltigkeit Berücksichtigung finden. In der vorliegenden Untersuchung erfolgt eine Konzentration auf die ökologische und soziale Dimension der Nachhaltigkeit. Dies ist im Wesentlichen zwei Überlegungen geschuldet. Zum einen liegen bereits zahlreiche Untersuchungen im Bereich der Finanzberichterstattung vor (für einen Überblick vgl. *Healy/Palepu* 2001, 405ff.). Zum anderen existieren gesetzliche Regelungen zur Bewertung und Berichterstattung ökonomischer Sachverhalte, so dass der Einbezug dieser Dimension die Ergebnisse einer Analyse des Zusammenhangs zwischen der freiwilligen Nachhaltigkeitsberichterstattung und -performance verzerren könnte.

3.2 Messung von Umfang und Form der Nachhaltigkeitsberichterstattung

In der vorliegenden Untersuchung wird *Nachhaltigkeitsberichterstattung* als Gesamtheit aller von einem Unternehmen der Öffentlichkeit absichtlich zugänglich gemachten Informationen über die ökonomischen, ökologischen und sozialen Aspekte seiner Entscheidungen abgegrenzt. Aus den in Abschnitt 3.1 genannten Gründen werden in dieser Arbeit die ökologischen und sozialen Aspekte herausgegriffen. In bisherigen empirischen Studien, die in *Tabelle 1* dargestellt sind, wird die Messung der Umweltberichterstattung überwiegend mittels Inhaltsanalyse vorgenommen. Hierbei wird die Berichterstattung zu definierten Informationsitems binär oder ordinal skaliert erfasst und die Einzelwerte werden zu einem Gesamtwert aggregiert (vgl. exemplarisch *Clarkson et al.* 2008 für eine binär skalierte und *Al-Tuwaijri et al.* 2004 für eine ordinal skalierte Messung). Ohne eine Unterscheidung der publizierten Informationen in quantitativ und verbal ist ein solches Messkonzept für die Beantwortung der Fragestellung nicht geeignet. Bei Binärcodierung spiegelt die Bewertung lediglich den Umfang der Berichterstattung wider und erlaubt demzufolge keine Aussagen zur Gültigkeit der Theorien. Selbst wenn eine präzise Unterscheidung in quantitativ/verbal auf Basis der Informationsitems vorgenommen wird, besteht die Gefahr, dass sich diese bei der Gesamtbewertung gegenseitig kompensieren. Diese Probleme können zwar durch eine in verbale Berichterstattung (gering) und quantitative Berichterstattung (hoch) ordinal trennende Bewertung überwunden werden, allerdings sind auch hier die Ergebnisse nicht eindeutig interpretierbar. So könnte ein positiver Zusammenhang sowohl als Beleg für die Gültigkeit der VDT herangezogen werden, wonach eine bessere Umweltpflege mit

einer vermehrt quantitativen Berichterstattung einhergeht, als auch als Beleg für die Gültigkeit der LT, wonach eine schlechtere Umweltperformance mit einer geringeren quantitativen und vermehrt verbalen Berichterstattung einhergeht.

Lediglich bei einigen Studien jüngeren Datums wird eine Unterscheidung von quantitativer und verbaler Form der Berichterstattung ansatzweise vorgenommen (vgl. Clarkson *et al.* 2008; Cho/Patten 2007; Al-Tuwarijri *et al.* 2004; Bewley/Li 2000). Ausgangspunkt für eine differenzierte Messung der Berichterstattung bei diesen Studien stellt die Zuordnung von Informationsitems zu einer harten/monetären *oder* einer weichen/nicht-monetären Berichterstattung dar. Zur Kategorie der weichen bzw. nicht-monetären Informationsitems werden in der Regel Angaben zu Strategie, Umweltinitiativen u.a. gezählt, wohingegen harte bzw. monetäre Informationsitems zentrale Umweltindikatoren wie Energieverbrauch oder Treibhausgasemissionen umfassen. Bei einem solchen Messansatz ist allerdings das Kriterium der inhaltlichen Kongruenz der Informationsitems nicht erfüllt. Sowohl bei Binärcodierung als auch bei einer ordinal skalierten Bewertung sind die Ergebnisse für die harte/monetäre Kategorie nicht eindeutig interpretierbar, da ein positiver Zusammenhang sowohl als Beleg für die Gültigkeit der LT (je schlechter die Umweltperformance, desto weniger harte Informationen werden berichtet) als auch als Beleg für die Gültigkeit der VDT (je besser die Umweltperformance, desto mehr harte Informationen werden berichtet) gedeutet werden kann. Dementsprechend fordern Clarkson *et al.* (2008, 325), „that future environmental disclosure research should move the focus of enquiry beyond the *level of disclosure*.“

Vor diesem Hintergrund soll im Folgenden ein Konzept zur Messung der Nachhaltigkeitsberichterstattung vorgestellt werden, welches für die Untersuchung der Problemstellung geeignet ist. Die in einem ersten Schritt erforderliche Auswahl der relevanten Informationselemente orientiert sich an den Leistungsindikatoren der aktuell gültigen Richtlinien der Global Reporting Initiative (GRI)² G3.1, welche weltweit quasi einen Standard der Nachhaltigkeitsberichterstattung darstellen (vgl. Hoffmann 2011, 68).³ Es werden jeweils zehn Einzelindikatoren für die Kategorien Ökologie und Soziales erfasst, wobei drei Leistungsindikatoren der Kategorie Ökologie für Unternehmen bestimmter Branchen durch branchenspezifische Angaben ersetzt werden. Zudem erfolgt je nach Bedeutung der Einzelindikatoren eine Gewichtung mit dem Faktor 1 oder 2. Tabelle 2 zeigt die berücksichtigten Leistungsindikatoren in den einzelnen Kategorien mit dem jeweiligen Bezug zu den Leistungsindikatoren der GRI-Richtlinien und dem Gewichtungsfaktor. Die mit * gekennzeichneten Informationselemente werden bei Unternehmen der entsprechenden Branche durch die angeführten branchenspezifischen Indikatoren ersetzt.

2 Bei der GRI handelt es sich um eine Nichtregierungsorganisation, welche im Jahre 1997 mit dem Ziel gegründet wurde, die weltweite Verbreitung und Standardisierung der Nachhaltigkeitsberichterstattung voranzutreiben. Vgl. *Global Reporting Initiative* (2012).

3 Diese Vorgehensweise wird auch in anderen Studien gewählt, vgl. beispielsweise Cho *et al.* (2012), Clarkson *et al.* (2011) oder Clarkson *et al.* (2008).

Abkürzung	Leistungsindikator	Bezug zu G3.1	Faktor
Ökologie			
R-B1	Eingesetztes Material nach Gewicht oder Volumen	EN1	2
R-B2	Anteil Recyclingmaterial am Gesamtmaterial einsatz	EN2	1
R-B3	Direkter und indirekter Energieverbrauch	EN3/4	2
R-B4	Gesamtwasserverbrauch nach Art	EN8	2
R-B5	Gesamte direkte und indirekte Treibhausgasemissionen	EN16	2
R-B6*	Emissionen von Ozon abbauenden Stoffen	EN19	1
R-B7	Andere Luf temissionen (Stickoxide, Schwefeloxide u.a.)	EN20	1
R-B8*	Gesamte Abwassereinleitungen nach Art und Einleitungsort	EN21	2
R-B9	Gesamtgewicht des Abfalls nach Art und Entsorgungs methode	EN22	2
R-B10*	Gesamtgewicht des gefährlich eingestuften Abfalls und dessen Entsorgung	EN22	1
Soziales			
R-C1	Anzahl Mitarbeiter nach mindestens zwei Kategorien (Region, Geschlecht, Qualifikation)	LA1	2
R-C2	Anzahl leitende Mitarbeiter nach mindestens zwei Kategorien (Region, Geschlecht, Qualifikation)	LA13	2
R-C3	Verhältnis Grundgehalt Männer/Frauen	LA14	2
R-C4	Mitarbeiterfluktuation	LA2	2
R-C5	Gesundheit: Arbeitsunfälle, Verletzungen, Ausfalltage, Todesfälle	LA7	2
R-C6	Anteil Mitarbeiter mit Kollektivvereinbarungen/ in Gewerkschaften	LA4	1
R-C7	Weiterbildungsumfang: zeitlicher oder finanzieller Aufwand	LA10	2
R-C8	Gesamtzahl der Vorfälle von Diskriminierung	HR4	1
R-C9	Verhinderung von Kinderarbeit: Umfang und Ergebnis von Audits	HR6	1
R-C10	Verhinderung von Zwangs- und Pflichtarbeit: Umfang und Ergebnis von Audits	HR7	1
			Summe 32
Branchenspezifische Leistungsindikatoren (ersetzen die mit * gekennzeichneten Leistungsindikatoren)			
Banken/Versicherungen			
R-D1	Vermögen in socially responsible investments (SRI) nach Kriterien	FS8/11	2
R-D2	Umfang Mikrofinanzierung/Mikroversicherung: Anzahl Kunden, investierte Mittel, Prämien	FS13/14	2
Handel			
R-D1	Kältemittel: Angaben zu Fluorkohlenwasserstoffen (FKW, H-FKW)	-	2
R-D2	Anteil fair gehandelter Produkte/Bioprodukte am Umsatz oder Sortiment	-	2
Transport			
R-D1	Flottenzusammensetzung im Hinblick auf CO ₂ -Effizienz	LT2	2
R-D2	Nutzung erneuerbarer Energiequellen zum Transport: Anteil Bio-Treibstoffe, Anzahl Hybrid-Fahrzeuge	LT3/4	2

Abkürzung	Leistungsindikator	Bezug zu G3.1	Faktor
Telekommunikation			
R-D1	Angaben zum Schutz von Kundendaten: Anzahl Verstösse, Anzahl und Ergebnis von Audits	PR8	2
R-D2	Anzahl der allein und gemeinsam genutzten Sendemasten	IO8	2
Software			
R-D1	Angaben zum Schutz von Kundendaten: Anzahl Verstösse, Anzahl und Ergebnis von Audits	PR8	2
R-D2	Angebote für gemeinnützige Organisationen: Anzahl oder Wert vergünstigter Software, Umfang konkreter Initiativen	-	2
Personaldienstleister			
-	Erhöhung der Gewichtungsfaktoren bei C4, C5, C6 und C7 um jeweils 1	-	-

Tabelle 2: Überprüfte Elemente der Nachhaltigkeitsberichterstattung mit Gewichtung

Sämtliche Informationselemente sind gemäss Vorgaben der GRI quantitativ zu erfassen. Die GRI gibt im Rahmen von Indikatorprotokollen ausserdem detaillierte Hinweise zur Berechnung der einzelnen Kennzahlen. Es wird geprüft, ob das jeweilige Informationselement berichtet wird (Umfang der Nachhaltigkeitsberichterstattung) und falls eine Berichterstattung vorliegt, ob diese quantitativ oder verbal ausfällt (Form der Nachhaltigkeitsberichterstattung). Als *verbale* Berichterstattung gelten Erläuterungen und Erörterungen sowie quantitative Informationen, aus denen die zu berichtende Kennzahl nicht ohne weitere Annahmen berechnet werden kann.⁴ Eine *quantitative* Berichterstattung impliziert die numerische Angabe der jeweiligen Kennzahl für das Berichterstattungsjahr für das Gesamtunternehmen. Auch wenn im Rahmen der Indikatorprotokolle zu den einzelnen Leistungsindikatoren mitunter mehrere zu berichtende Kennzahlen aufgelistet sind, wird im Rahmen dieses Messkonzepts bereits die Berichterstattung einer dieser Kennzahlen für das jeweilige Informationselement als „quantitative“ Berichterstattung bewertet (vgl. exemplarisch R-C5 mit Bezug zu LA7). Die gewichteten Bewertungen der Informationsitems werden zu einer Gesamtbewertung aufsummiert, wobei sowohl der Umfang als auch das quantitative/verbale Ausmass der Nachhaltigkeitsberichterstattung jeweils auf einer Skala von 0 (niedrigste Ausprägung) bis 32 (höchste Ausprägung, entspricht Summe aller Gewichtungsfaktoren) gemessen wird.

Das dargestellte Messkonzept ist geeignet, Umfang (1. Analysestufe) und quantitatives/verbales Ausmass (2. Analysestufe) der Nachhaltigkeitsberichterstattung zu bestimmen und damit die zentralen Forschungsfragen dieser Untersuchung zu beantworten. Da die Unterscheidung der Informationsart in quantitativ/verbal an der Bewertung und nicht am Informationselement anknüpft, können die Probleme der Messkonzepte bisheriger Studien weitestgehend überwunden werden. Allerdings bedingt die Fokussierung auf quantitative Leistungsindikatoren eine überschaubare Anzahl an Indikatoren, die jedoch vergleichbar zu bisheriger Studien ausfällt (mit Ausnahme der Untersuchung von Clarkson *et al.* 2008).

4 Beispiele hierfür sind die numerische Berichterstattung für lediglich ein ausgewähltes Produkt oder einen ausgewählten Teilbereich des Unternehmens oder die ausschliessliche Angabe der prozentualen Veränderung des Leistungsindikators gegenüber dem Vorjahr.

Durch eine detaillierte Beschreibung der Leistungsindikatoren und Bewertungsweise können Objektivität und Reliabilität der Messung sichergestellt werden. Auch die Validität des Messansatzes erscheint im Hinblick auf die Zielseitung erfüllt. Dem möglichen Einwand, wonach das vorgestellte Messkonzept keine vollständige Bewertung der Nachhaltigkeitsberichterstattung wie bei anderen Studien erlaubt (vgl. exemplarisch *Quick/Knocinski 2006; Morhardt et al. 2002*), muss entgegengestellt werden, dass nicht alle denkbaren Berichtselemente sinnvoll in quantitativ und verbal getrennt werden können. Dementsprechend erscheint die Konzentration auf ausgewählte, relevante Leistungsindikatoren der Nachhaltigkeitsberichterstattung gerechtfertigt. Durch die Berücksichtigung der zwei Dimensionen Ökologie und Soziales wird weiterhin dem Kriterium der Mehrdimensionalität Rechnung getragen. Der folgende Abschnitt zeigt, dass auch die Anforderung der Kongruenz der beiden Messkonzepte erfüllt ist.

3.3 Messung der Nachhaltigkeitsperformance

In Anlehnung an eine Definition von Performance im Management-Kontext (vgl. *Breuer et al. 2010, 2328*) wird die *Nachhaltigkeitsperformance* eines Unternehmens definiert als seine tatsächliche Leistung im Hinblick auf ökonomische, ökologische und soziale Aspekte. Für die vorliegende Untersuchung erfolgt aus den in Abschnitt 3.1 genannten Gründen eine Konzentration auf die ökologischen und sozialen Aspekte der Nachhaltigkeitsperformance. Die Beurteilung der Performance erfordert die Messbarkeit sowie Vergleichbarkeit mittels geeigneter Bezugspunkte. Aufgrund der aufgezeigten Schwächen bisheriger Messansätze sind diese als Grundlage für die Entwicklung eines neuen Messkonzepts nicht geeignet. Vielmehr wird ein iteratives Vorgehen gewählt, dessen Ausgangspunkt entsprechend des in Abschnitt 3.1 angeführten Kriteriums der Kongruenz die Informationsitems des Messkonzepts der Nachhaltigkeitsberichterstattung darstellen.

In einem ersten Schritt wurde die Berichterstattungspraxis sämtlicher Unternehmen des Untersuchungssamples zu diesen Informationsitems analysiert. Dabei zeigte sich, dass zahlreiche der in Abschnitt 3.2 berücksichtigten Leistungsindikatoren überhaupt nicht oder nicht in quantitativer Form berichtet werden. Des Weiteren berichten die Unternehmen bezüglich eines Leistungsindikators mitunter *verschiedene* quantitativ messbare Sachverhalte (vgl. Leistungsindikator R-C5). Der vorliegende Ansatz zur Messung der Nachhaltigkeitsperformance berücksichtigt diese Problematik, indem eine Konzentration auf jene Leistungsindikatoren des Messkonzepts zur Bewertung der Nachhaltigkeitsberichterstattung erfolgt, zu denen die meisten Unternehmen des Untersuchungssamples quantitative Angaben vornehmen. Die nachfolgende Tabelle gibt einen Überblick über die Leistungsindikatoren des Messkonzepts der Nachhaltigkeitsperformance. Für jeden Leistungsindikator werden Berechnung und Masseinheit sowie der Bezug zum Messkonzept aus Abschnitt 3.2 aufgezeigt. Insgesamt werden acht Leistungsindikatoren erhoben, davon jeweils vier für die Dimensionen Ökologie und Soziales.

Artikel

Abkürzung	Leistungsindikator	Berechnung	Masseneinheit	Bezug Berichterstattung
Ökologie				
P-B1	Energieverbrauch	(direkter + indirekter Energieverbrauch) / Anzahl MA	MWh / MA	R-B3
P-B2	Wasserentnahme	(gesamte Wasserentnahme – Kühlwasser) / Anzahl MA	m³ / MA	R-B4
P-B3	Treibhausgasemissionen	(direkte + indirekte Emissionen in CO ₂ -Äquivalenten) / Anzahl MA	t / MA	R-B5
P-B4	Abfallmenge	Gesamtgewicht des Abfalls / Anzahl MA	t / MA	R-B9
Soziales				
P-C1	Mitarbeiterfluktuation	Anzahl aller Austritte / Anzahl MA	Prozent	R-C4
P-C2	Weiterbildung	Gesamte Weiterbildungszeit / Anzahl MA	h / MA	R-C7
P-C3	Frauenanteil in den obersten Exekutiv- und Kontrollorganen	(Anzahl Frauen in VS + AR bzw. GL + VR) / Anzahl Mitglieder bei den Gremien	Prozent	R-C2
P-C4	Verantwortungsgerechte Vergütung des obersten Exekutivorgans	(Gesamtvergütung VS bzw. GL / Anzahl der Mitglieder) / Anzahl MA	€ / MA	-

AR Aufsichtsrat GL Geschäftsleitung MA Mitarbeiter in Vollzeitäquivalenten

VR Verwaltungsrat VS Vorstand

Tabelle 3: Indikatoren zur Messung der Nachhaltigkeitsperformance

Es wird deutlich, dass gemäss dem Kriterium der Mehrdimensionalität die wesentlichen Aspekte der beiden analysierten Dimensionen der Nachhaltigkeit jeweils Berücksichtigung finden. Ausserdem zeigt ein Vergleich der Leistungsindikatoren der beiden Messkonzepte, dass auch das Kriterium der Kongruenz weitestgehend erfüllt ist. Lediglich für Indikator P-C4 gibt es keine direkte Entsprechung zur Messung der Nachhaltigkeitsberichterstattung, da aufgrund gesetzlicher Offenlegungspflichten eine Erfassung dieses Leistungsindikators im Rahmen der Nachhaltigkeitsberichterstattung nicht zweckmässig ist (vgl. Art. 663b und Art. 663c Abs. 3 OR sowie § 314 Abs. 1 Nr. 6a HGB).

Nachfolgend werden Besonderheiten der verwendeten Leistungsindikatoren kurz erläutert. Die Leistungsindikatoren P-B1 bis P-B4 stellen klassische umweltbezogene Leistungsindikatoren zu Energieverbrauch, Wasserentnahme, Treibhausgasemissionen sowie zur Abfallmenge dar. Bei sämtlichen Leistungsindikatoren erfolgt eine Berücksichtigung der Unternehmensgrösse durch Relativierung der Kennzahl über die Gesamtzahl an beschäftigten Mitarbeitern (in Vollzeitäquivalenten). Bei den Treibhausgasemissionen wird sofern vorhanden ein Abgleich mit den Angaben des Carbon Disclosure Projects vorgenommen. Es werden die tatsächlich emittierten Treibhausgase ohne Einbezug einer Kompensation durch Emissionszertifikate berücksichtigt. Die Leistungsindikatoren P-C1 bis P-C3 spie-

geln die unternehmensinterne, d.h. mitarbeiterbezogene Perspektive der sozialen Dimension wider, Leistungsindikator P-C4 die unternehmensexterne, d.h. gesellschaftliche Perspektive. Leistungsindikator P-C3, welcher einen Aspekt der Chancengleichheit von Frauen und Männern misst, berücksichtigt lediglich den Anteil von Frauen in Geschäftsleitung und Verwaltungsrat bzw. in Vorstand und Aufsichtsrat. Weitere Indikatoren, wie das Gehaltsverhältnis von Frauen und Männern auf gleicher Qualifikationsstufe bzw. der Anteil von Frauen in Leitungspositionen im gesamten Unternehmen, bleiben aufgrund fehlender oder nicht vergleichbarer Angaben der einzelnen Unternehmen unberücksichtigt. Leistungsindikator P-C4 misst die durchschnittliche Vergütung eines Geschäftsleitungs- bzw. Vorstandsmitglieds je Mitarbeiter und dient als Indikator für die gesellschaftliche Akzeptanz der Geschäftsleitungs- bzw. Vorstandsvergütung. Analog zu den Leistungsindikatoren P-B1 bis P-B4 erfolgt auch hier eine Berücksichtigung der Unternehmensgröße durch den Einbezug der Mitarbeiterzahl im Nenner. Diese Relativierung soll dem Verantwortungsumfang der Geschäftsleitung, welcher sich auch in einer höheren Vergütung ausdrückt, Rechnung tragen. Auch wenn die absolute Höhe der durchschnittlichen Vergütung von Geschäftsleitung bzw. Vorstand je Mitarbeiter keine direkten Rückschlüsse auf ihre gesellschaftliche Akzeptanz zulässt, so kann dennoch vermutet werden, dass die gesellschaftliche Akzeptanz mit zunehmender Vergütungshöhe sinkt und diese somit als Indikator für die soziale Dimension der Nachhaltigkeitsperformance eines Unternehmens herangezogen werden kann.

Analog zu Abschnitt 3.2 werden die acht Leistungsindikatoren zu einem Gesamtwert verdichtet. In einem ersten Schritt erfolgt eine möglichst präzise und einheitliche quantitative Erfassung der einzelnen Leistungsindikatoren gemäss den Angaben zur Berechnung in *Tabelle 3*. Hierfür werden die von den Unternehmen im Rahmen der Nachhaltigkeitsberichterstattung sowie der Geschäftsberichterstattung publizierten Informationen manuell ausgewertet. Die möglichst präzise Datenerfassung ist mit zahlreichen Schwierigkeiten aufgrund von ungenauen, unvollständigen oder uneinheitlichen Angaben bei der Berichterstattung der Unternehmen verbunden. Soweit möglich werden direkt die jeweiligen Angaben des Unternehmens übernommen. Falls die Informationen lediglich für ausgewählte Bereiche, Regionen oder Mitarbeitergruppen ausgewiesen sind, werden die vorhandenen Daten auf das Gesamtunternehmen hochgerechnet. Dabei lassen sich Verzerrungen aufgrund der teilweise unvollständigen Berichterstattung nicht ausschliessen. Sofern überhaupt keine Angaben vorliegen, enthält der Datensatz an dieser Stelle fehlende Werte. Ist dies für mehr als zwei Indikatoren je Dimension der Fall, wird der Datensatz komplett ausgeschlossen. In einem zweiten Schritt werden die erhobenen Daten auf eine Skala [0, 100] transformiert. Hierfür werden die analysierten Unternehmen gemäss ihrer Branche in die Klassen Dienstleistung, Industrie und Energieversorgung eingeteilt. Insbesondere im Hinblick auf die ökologische Dimension soll hierdurch sichergestellt werden, dass die Bewertung den branchenüblichen Gegebenheiten Rechnung trägt. In jeder Klasse werden der „schlechtesten“ Ausprägung eines Leistungsindikators (z.B. beim Energieverbrauch dem höchsten Wert, bei der Weiterbildung dem niedrigsten Wert) null und der „besten“ Ausprägung 100 Bewertungspunkte zugewiesen. Die dazwischen liegenden Werte werden proportional reskaliert. Durch Bildung des arithmetischen Mittelwerts der Bewertungen der einzelnen Leistungsindikatoren erhält man die durchschnittliche Bewertung der zwei Dimensionen. Fehlende Werte werden so durch den Mittelwert der übrigen Indikatoren der jeweiligen Dimension approximiert. Das arithmetische Mittel der Bewertung der beiden Dimensionen

ergibt wiederum den Gesamtwert der Nachhaltigkeitsperformance, welcher auf einer theoretischen Skala [0, 100] liegt. Die getrennte Transformation der Originalwerte für die einzelnen Branchenklassen soll den stark unterschiedlichen Ausprägungen in den jeweiligen Klassen Rechnung tragen. Dies gilt insbesondere für die Dimension Ökologie, jedoch muss auch für die Dimension Soziales die tendenziell höhere Bedeutung des Humankapitals für Dienstleistungsunternehmen berücksichtigt werden. Die Berücksichtigung von Branchenklassen erfordert eine hinreichend grosse Anzahl an Unternehmen in den jeweiligen Klassen. Grundsätzlich ist auch eine branchenspezifischere Unterscheidung denkbar, um eine bessere Vergleichbarkeit der Daten innerhalb dieser Klassen zu gewährleisten. Allerdings muss berücksichtigt werden, dass auch bei einer engen branchenspezifischen Klasseneinteilung andersartige Geschäftsmodelle der jeweiligen Unternehmen unterschiedliche Performancedaten implizieren. Eine direkte Vergleichbarkeit der Daten wäre auch dann nicht ohne Weiteres gegeben.

Trotz dieser Einschränkungen ist das vorgestellte Messkonzept geeignet, die aufgezeigten Schwächen bisheriger Ansätze im Hinblick auf die in Abschnitt 3.1 angeführten Gütekriterien weitestgehend zu überwinden. Durch die Anknüpfung des Messkonzepts an zentrale ökologische und soziale Leistungsindikatoren kann die Objektivität der Messung sichergestellt werden. Dies wirkt sich auch positiv auf die Reliabilität der Messung aus, welche lediglich durch die Reskalierung der Daten auf Basis der minimalen und maximalen Ausprägung beeinträchtigt sein könnte. Auch die Validität ist im Vergleich zu bisherigen Messkonzepten als weitestgehend erfüllt zu beurteilen. Zwar ist eine vollständige Erfassung sämtlicher ökologischer und sozialer Auswirkungen wünschenswert, dies ist jedoch aufgrund der Berichterstattungspraxis nicht möglich. Diese Problematik wird durch eine hohe Relevanz der berücksichtigten Leistungsindikatoren gemildert. Im Hinblick auf die Reskalierung der Originaldaten muss einschränkend auf die relativ grobe Unterscheidung von Branchenklassen hingewiesen werden. Weitestgehend erfüllt sind die Kriterien Kongruenz und Mehrdimensionalität, da wesentliche Elemente der Nachhaltigkeitsberichterstattung auch bei der Performancemessung in den Bereichen Ökologie und Soziales berücksichtigt werden.

Unabhängig von der Erfüllung der Gütekriterien ist das vorgestellte Messkonzept lediglich im Hinblick auf die Anlehnung an unternehmensseitig zur Verfügung gestellte Daten und eine damit einhergehende mangelnde Überprüfbarkeit kritisch zu beurteilen. Eine Verzerrung der Ergebnisse zur Nachhaltigkeitsperformance durch fehlende oder falsche Angaben der berichtenden Unternehmen ist nicht auszuschliessen. Dieser Effekt wird durch den Ausschluss von Unternehmen, welche keine Informationen zur Nachhaltigkeitsperformance ihres Unternehmens veröffentlichen, verstärkt. Im Hinblick auf Unternehmen, die einen oder zwei fehlende Werte je Dimension aufweisen und daher noch berücksichtigt werden, ist die Verzerrung für die Hypothesenprüfung jedoch unkritisch. Sofern Unternehmen mit schlechter Performance tatsächlich mehr fehlende Werte aufweisen, wird die Performance durch die Mittelwertbildung über vorhandene Daten tendenziell überschätzt. Dies schwächt die im vierten Abschnitt ermittelten Zusammenhänge ab, es bewirkt jedoch gerade keine Verzerrung in Richtung der Bestätigung unserer Hypothesen. Zukünftig könnte die Problematik infolge einer zunehmenden Verbreitung der externen Überprüfung (Assurance) von im Rahmen der Nachhaltigkeitsberichterstattung publizierten Informationen (vgl. Kolk 2010) an Bedeutung verlieren. Insgesamt stellt das Messkonzept daher im Vergleich zu den Ansätzen bisheriger Studien eine deutliche Verbesserung dar.

4. Empirische Ergebnisse

4.1 Bestimmung des Untersuchungssamples

Die vorliegende Studie basiert auf einer Analyse der Nachhaltigkeitsberichterstattung von 50 Deutschen und Schweizer Grossunternehmen. Neben reinen Nachhaltigkeitsberichten werden auch Geschäfts-, Umwelt- und Personalberichte sowie andere themenbezogene Online-Publikationen einbezogen. Diese Vorgehensweise berücksichtigt, dass noch nicht alle Unternehmen ihre Nachhaltigkeitsberichterstattung in einem Dokument bündeln. Die Grundgesamtheit umfasst sämtliche Unternehmen, welche zum 31.12.2011 im Deutschen Aktienindex (DAX) oder im Swiss Leader Index (SLI) kotiert sind. Von der Analyse ausgeschlossen werden Unternehmen der Energieversorgungsbranche, da die Stichprobe lediglich zwei Unternehmen dieser Branche (E.ON und RWE) beinhaltet. Diese Klassengrösse ist für eine aussagekräftige Reskalierung der erhobenen Daten zur Nachhaltigkeitsperformance nicht ausreichend. Ergänzend wird die Nachhaltigkeitsberichterstattung und -performance der Detailhändler Migros und Coop analysiert, da diese im Hinblick auf die öffentliche Wahrnehmung eine besondere Bedeutung in der Schweiz innehaben. Der Analysezeitraum umfasst das Berichterstattungsjahr 2010. Für zehn Unternehmen (Actelion, Nobel Biocare, Fresenius Medical Care, Fresenius SE, Infineon, Logitech, Swatch, Synthes, Transocean sowie Zurich) liegen keine Informationen zur Nachhaltigkeit in auswertbarem Umfang vor, weshalb sich die Grösse der Stichprobe auf 50 Unternehmen reduziert. Von diesen können 19 Unternehmen der Dienstleistungsbranche (inklusive Handel) und 31 Unternehmen der Industriebranche (inklusive Transport) zugeordnet werden.

Die Konzentration auf Deutsche und Schweizer Grossunternehmen trägt mehreren Überlegungen Rechnung. So zeigt eine Reihe von empirischen Untersuchungen einen signifikant positiven Einfluss der Unternehmensgrösse auf den Umfang bzw. die Qualität der Nachhaltigkeitsberichterstattung (vgl. Clarkson *et al.* 2011; Clarkson *et al.* 2008; Patten 2002; Bewley/Li 2000). Dieser Effekt dürfte bei einem Untersuchungssample mit hinreichend grossen Unternehmen weniger stark ausfallen. Zudem ist hierdurch sichergestellt, dass unternehmensintern ausreichend Ressourcen zur Erstellung eines Nachhaltigkeitsberichts und zur aktiven Steuerung der Nachhaltigkeitsperformance zur Verfügung stehen. Des Weiteren erzielen Grossunternehmen in der Regel einen relativ hohen Umsatz und beschäftigen eine grosse Anzahl an Mitarbeitern, so dass sie daran gemessen einen hohen Anteil der gesamten Unternehmenslandschaft abbilden. Da es bislang keine empirische Untersuchung des Zusammenhangs zwischen Nachhaltigkeitsperformance und -berichterstattung für europäische Unternehmen gibt, erscheint eine Konzentration auf Deutsche und Schweizer Grossunternehmen in einer ersten Studie gerechtfertigt. Letztlich führt die aufwändige Form der Datenerhebung zu einer Begrenzung des Stichprobenumfangs. Sämtliche Daten zur Nachhaltigkeitsberichterstattung und -performance wurden von den Verfassern manuell erhoben.

4.2 Deskriptive Ergebnisse

Die deskriptiven Ergebnisse zu Umfang und Form der Nachhaltigkeitsberichterstattung sind in *Tabelle 4* dargestellt. Für jeden Leistungsindikator des Messkonzepts – mit Ausnahme der branchenspezifischen Leistungsindikatoren – ist der Anteil an Unternehmen getrennt nach Industrie- und Dienstleistungsbranche dargestellt, der diesen nicht, verbal oder quantitativ berichtet.

	Industrieunternehmen				Dienstleistungsunternehmen			
	n	nicht berichtet	verbal berichtet	quantitativ berichtet	n	nicht berichtet	verbal berichtet	quantitativ berichtet
R-B1	31	32 %	32 %	35 %	19	26 %	32 %	42 %
R-B2	31	42 %	48 %	10 %	19	42 %	16 %	42 %
R-B3	31	0 %	26 %	74 %	19	0 %	26 %	74 %
R-B4	31	10 %	35 %	55 %	19	0 %	21 %	79 %
R-B5	31	0 %	26 %	74 %	19	0 %	11 %	89 %
R-B6*	28	25 %	14 %	61 %	-	-	-	-
R-B7	31	19 %	10 %	71 %	19	84 %	5 %	11 %
R-B8*	28	21 %	21 %	57 %	-	-	-	-
R-B9	31	6 %	29 %	65 %	19	11 %	47 %	42 %
R-B10*	28	29 %	7 %	64 %	-	-	-	-
Mittelwert		18 %	25 %	57 %		23 %	23 %	54 %
R-C1	31	0 %	19 %	81 %	19	0 %	11 %	89 %
R-C2	31	16 %	52 %	32 %	19	21 %	79 %	0 %
R-C3	31	55 %	45 %	0 %	19	47 %	47 %	5 %
R-C4	31	23 %	19 %	58 %	19	16 %	16 %	68 %
R-C5	31	10 %	16 %	74 %	19	11 %	63 %	26 %
R-C6	31	13 %	45 %	42 %	19	21 %	32 %	47 %
R-C7	31	10 %	32 %	58 %	19	16 %	37 %	47 %
R-C8	31	23 %	52 %	26 %	19	26 %	47 %	26 %
R-C9	31	6 %	84 %	10 %	19	5 %	58 %	37 %
R-C10	31	10 %	81 %	10 %	19	11 %	63 %	26 %
Mittelwert		17 %	45 %	39 %		17 %	45 %	37 %

Tabelle 4: Deskriptive Ergebnisse zu Umfang und Form der Nachhaltigkeitsberichterstattung

Die Ergebnisse zeigen, dass insbesondere die ökologischen Leistungsindikatoren von einem hohen Anteil an Unternehmen quantitativ berichtet werden. Im Durchschnitt etwas geringer fällt der Anteil an Unternehmen mit quantitativer Berichterstattung bei den sozialen Leistungsindikatoren aus. Für einige Indikatoren (R-B2, R-C3, R-C9 und R-C10) wird deutlich, dass diese von Industrieunternehmen überhaupt nicht oder überwiegend verbal berichtet werden. So wird beispielsweise von keinem Unternehmen dieser Gruppe eine konkrete quantitative Angabe zum Verhältnis des Grundgehalts zwischen Männern und Frauen (R-C3) gemacht. Obwohl diese Thematik sowohl in Deutschland als auch in der Schweiz von den Medien intensiv diskutiert wird, unterlassen die Unternehmen an dieser Stelle eine quantitative Berichterstattung. Von den Dienstleistungsunternehmen werden hingegen zu den Indikatoren R-B7, R-C2 und R-C3 nur vereinzelt quantitative Informationen berichtet. Während dies bei R-B7 (andere Luftemissionen) vermutlich der geringen Relevanz des Indikators zugeschrieben werden kann, deutet der Verzicht auf quantitative Angaben bei R-C2 (Anzahl leitende Mitarbeiter nach Kategorien) eher auf eine gezielt zurückhaltende Informationspolitik hin. Ähnlich zu den Industrieunternehmen berichten auch die beobachteten Dienstleistungsunternehmen kaum über das Verhältnis des Grundgehalts zwischen Männern und Frauen (R-C3).

In Tabelle 5 sind die wesentlichen deskriptiven Merkmale der Leistungsindikatoren sowie der Gesamtbewertung der Nachhaltigkeitsperformance jeweils getrennt für Industrie- und Dienstleistungsunternehmen zusammengefasst.

	P-B1	P-B2	P-B3	P-B4	P-C1	P-C2	P-C3	P-C4	Score
Industrieunternehmen									
n	31	26	31	25	19	18	31	31	31
MW	211,59	940,61	119,02	5,19	0,08	17,59	0,09	61,53	71,16
MD	45,70	300,50	14,32	3,07	0,09	18,14	0,09	42,42	72,39
SD	363,07	2.078,77	275,01	5,82	0,05	6,57	0,06	55,65	13,40
MIN	3,60	8,54	1,75	0,15	0,01	2,02	0	4,08	41,74
MAX	1.583,33	10.498,00	1.263,85	22,18	0,22	30,3	0,21	197,59	90,73
Dienstleistungsunternehmen									
n	19	19	19	18	17	14	19	19	19
MW	16,23	43,82	4,39	0,72	0,12	23,64	0,11	137,37	64,41
MD	14,78	18,90	3,98	0,30	0,11	18,00	0,09	57,87	63,84
SD	9,30	69,15	2,85	0,99	0,05	18,23	0,06	206,41	11,99
MIN	3,15	10,06	1,58	0,06	0,05	6,40	0,04	7,29	39,24
MAX	32,81	309,09	12,80	3,38	0,25	80,00	0,25	724,32	83,84

MW Mittelwert MD Median SD Standardabweichung MIN Minimum MAX Maximum

Tabelle 5: Deskriptive Ergebnisse zur Nachhaltigkeitsperformance

Die Ergebnisse zeigen teilweise deutliche Unterschiede bei den Leistungsindikatoren für Industrie- und Dienstleistungsunternehmen. Während die ökologischen Leistungsindikatoren wie erwartet bei Industrieunternehmen im Durchschnitt schlechter ausfallen als bei Dienstleistungsunternehmen, erhält man für die soziale Dimension ein weniger eindeutiges Ergebnis. Hier schneiden Industrieunternehmen bei der Mitarbeiterfluktuation (P-C1) sowie bei der verantwortungsgerechten Vergütung des obersten Exekutivorgans (P-C4) besser und bei den übrigen beiden Leistungsindikatoren tendenziell schlechter ab als Dienstleistungsunternehmen. Insgesamt erscheint die getrennte Reskalierung der Leistungsindikatoren daher insbesondere im Hinblick auf die ökologische Dimension zweckmäßig, um die unterschiedlichen Streuungen der Performancedaten bei Industrie- und Dienstleistungsunternehmen zu berücksichtigen. Allerdings weisen das deutliche Auseinanderfallen von arithmetischem Mittelwert und Median sowie die grosse Streuung der Ergebnisse bei den Industrieunternehmen darauf hin, dass für diese Gruppe eine weitere Unterteilung angebracht sein könnte.

4.3 Hypothesenprüfung

Für die Hypothesenprüfung wird auf Pearson-Korrelationsanalysen zurückgegriffen.⁵ Eine Überprüfung von Hypothese H₁ ergibt einen positiven und signifikanten Zusammenhang ($r=0,33$; $p=0,02$; $n=50$) zwischen der Nachhaltigkeitsperformance und dem Umfang der

⁵ Die Ergebnisse des Shapiro-Wilk-Tests auf Normalverteilung indizieren mit einer Ausnahme für sämtliche zu untersuchenden Variablen keine Ablehnung der Normalverteilungshypothese bei einem Signifikanzniveau von 10%.

Nachhaltigkeitsberichterstattung. In *Abbildung 1* ist dieser Zusammenhang – unterschieden nach Industrie- und Dienstleistungsunternehmen – grafisch dargestellt.

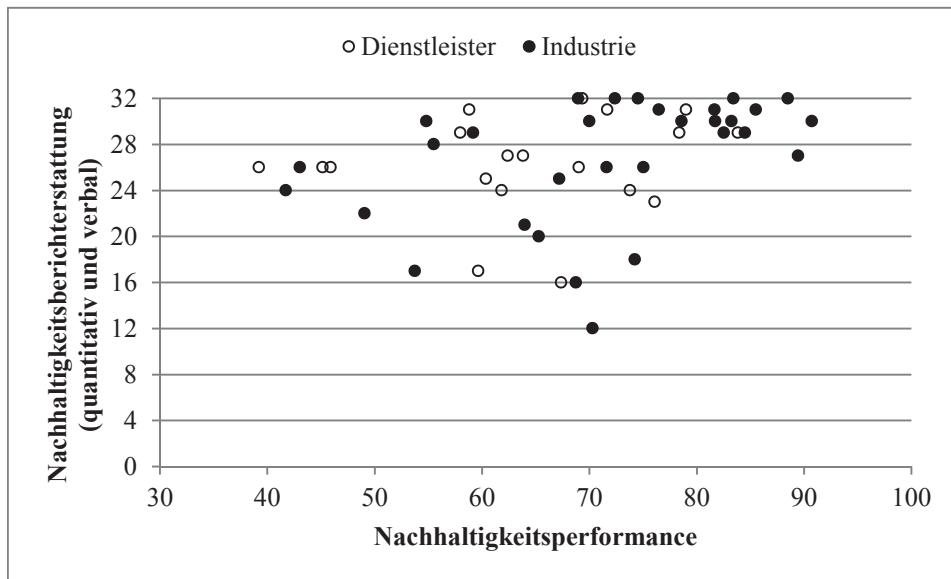
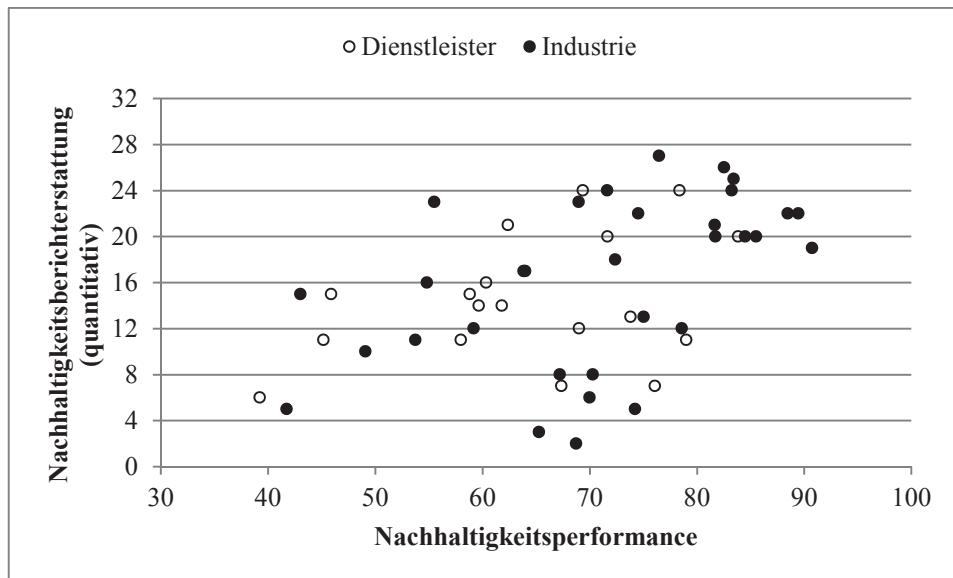
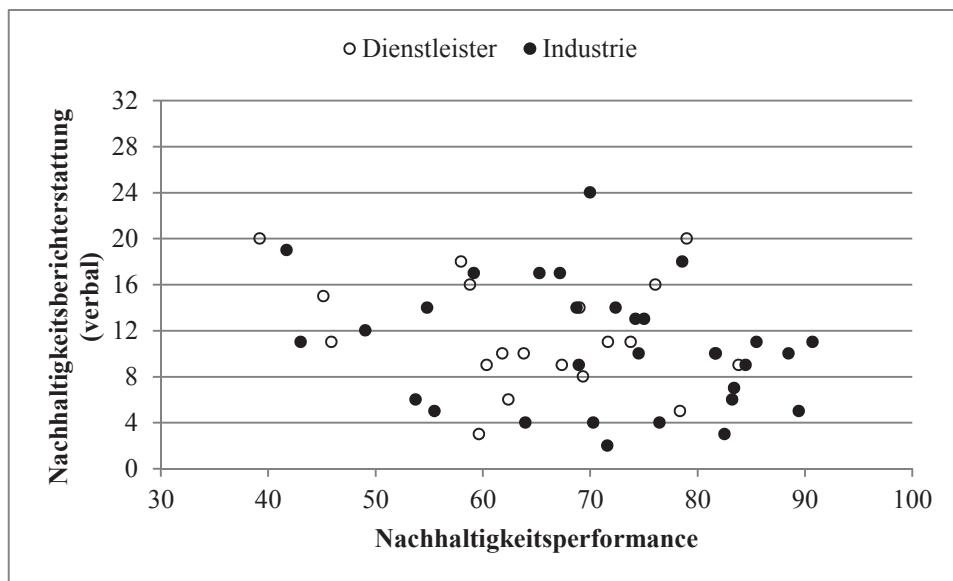


Abbildung 1: Grafische Zusammenhangsanalyse zu Hypothese H₁

Auch die Ergebnisse zur quantitativen Form der Nachhaltigkeitsberichterstattung stehen im Einklang mit unserer Argumentation. Der Zusammenhang zwischen Nachhaltigkeitsperformance und dem quantitativen Ausmass der Nachhaltigkeitsberichterstattung ist positiv und signifikant ($r=0,46$; $p=0,00$; $n=50$). Eine grafische Veranschaulichung mittels Streupunktdiagramm findet sich in *Abbildung 2*.

Abbildung 2: Grafische Zusammenhangsanalyse zu Hypothese H_{2a}Abbildung 3: Grafische Zusammenhangsanalyse zu Hypothese H_{2b}

Je besser die Nachhaltigkeitsperformance, desto eher veröffentlichen Unternehmen quantitative Informationen im Rahmen der Nachhaltigkeitsberichterstattung. Beide Hypothesen, welche einen Teil der Nachhaltigkeitsberichterstattung auf Basis der VDT erklären, werden somit gestützt.

Im Rahmen von *Hypothese H_{2b}* wird der Argumentation der LT folgend ein negativer Zusammenhang zwischen Nachhaltigkeitsperformance und verbaler Form der Nachhaltigkeitsberichterstattung postuliert. Auch diese Hypothese wird durch die empirischen Ergebnisse gestützt. Es besteht ein negativer und auf dem 10%-Niveau signifikanter Zusammenhang ($r=-0,27$; $p=0,06$; $n=50$) zwischen der Nachhaltigkeitsperformance und dem Ausmass der verbalen Nachhaltigkeitsberichterstattung (vgl. Abbildung 3).

Insgesamt bestätigen somit die empirischen Ergebnisse sämtliche Hypothesen der Untersuchung. Nachfolgend soll nun mittels Sensitivitätsanalysen die Robustheit der Ergebnisse untersucht werden.

Hierfür wird in einem ersten Schritt der Ansatz zur Messung der Nachhaltigkeitsberichterstattung im Hinblick auf die Wahl der Gewichtungsfaktoren analysiert. Es wird untersucht, inwiefern die Gewichtung der einzelnen Berichtselemente die Untersuchungsergebnisse beeinflusst. Neu fliessen sämtliche Berichtselemente mit einem Gewichtungsfaktor von 1 in das alternative Bewertungskonzept ein.⁶ In diesem alternativen Modell fallen die Ergebnisse hinsichtlich Richtung und Signifikanz nahezu identisch zu unserem Ausgangsmodell aus (H_1 : $r=0,33$; $p=0,02$; $n=50$. H_{2a} : $r=0,47$; $p=0,00$; $n=50$. H_{2b} : $r=-0,26$; $p=0,07$; $n=50$). Somit wird deutlich, dass unsere Untersuchungsergebnisse nicht durch die Wahl der Gewichtung getrieben werden.

In einem zweiten Schritt soll der Ansatz zur Messung der Nachhaltigkeitsperformance untersucht werden. In Anlehnung an die Vorgehensweise bisheriger Studien (vgl. Cho/Patten 2007) wird für die statistischen Auswertungen auf eine binäre Skalierung der Nachhaltigkeitsperformance zurückgegriffen. Hierfür wird das intervallskalierte Mass der Nachhaltigkeitsperformance durch eine Unterscheidung in high/low performer (über/unter Median) ersetzt. Die Ergebnisse eines Mittelwertvergleichs stützen erneut die *Hypothesen H₁* und *H_{2a}*, jedoch fällt das Ergebnis für *Hypothese H_{2b}* nicht signifikant aus.⁷ Die Analyse zeigt somit, dass eine intervallskalierte Messung der Nachhaltigkeitsperformance einer binären Unterscheidung überlegen ist. Des Weiteren wird die Problematik fehlender Werte bei einzelnen Leistungsindikatoren untersucht. In unserem Grundmodell werden diese fehlenden Werte durch den Mittelwert der übrigen Indikatoren in der jeweiligen Dimension ersetzt. Diese Vorgehensweise könnte eine Abschwächung der untersuchten Effekte bewirken, da gemäss VDT anzunehmen ist, dass vor allem schlechte Performancewerte nicht berichtet werden. Andererseits kann insbesondere für die ökologische Dimension von einem ähnlichen Niveau der Leistungsindikatoren innerhalb eines Unternehmens ausgegangen werden, so dass der Mittelwert der übrigen Performancewerte eine zumindest akzeptable Approximation des tatsächlichen Wertes darstellen könnte. Sämtliche Analysen werden erneut ausschliesslich für vollständige Datensätze durchgeführt. Die Ergebnisse fallen bezüglich Richtung und Signifikanz vergleichbar zu unserem Ausgangsmodell aus (H_1 : $r=0,38$; $p=0,07$; $n=24$. H_{2a} : $r=0,56$; $p=0,00$; $n=24$. H_{2b} : $r=-0,34$; $p=0,11$; $n=24$). Lediglich die Ergebnisse für *Hypothese H_{2b}* sind in diesem Modell knapp nicht signifikant zum 10%-Niveau. Hierbei ist jedoch der mit $n=24$ geringere Stichprobenumfang zu beachten. Insges-

6 Davon ausgenommen sind die mit * gekennzeichneten Leistungsindikatoren, bei welchen die Gewichtung unverändert erhalten bleibt. Die Substitution mit branchenspezifischen Leistungsindikatoren würde andernfalls zu unterschiedlichen maximal möglichen Gesamtbewertungen für die einzelnen Unternehmen führen.

7 Die Ergebnisse lauten: H_1 : MW(low)=24,88; MW(high)=28,20; $p(T<t)=0,01$. H_{2a} : MW(low)=13,12; MW(high)=17,96; $p(T<t)=0,00$. H_{2b} : MW(low)=11,76; MW(high)=10,24; $p(T>t)=0,15$.

samt scheint daher der Umgang mit fehlenden Werten in unserem Messkonzept der Nachhaltigkeitsperformance zu keiner systematischen Verzerrung der Ergebnisse zu führen.

In einem letzten Schritt wird die Robustheit der Ergebnisse bei Berücksichtigung von Kontrollvariablen untersucht. Den Ergebnissen bisheriger Untersuchungen folgend wird für die Grösse der untersuchten Unternehmen kontrolliert (vgl. Clarkson *et al.* 2011; Clarkson *et al.* 2008; Patten 2002; Bewley/Li 2000). Die Unternehmensgrösse wird über die Mitarbeiterzahl operationalisiert. Für die Analyse erfolgt basierend auf dem Median eine Unterscheidung in kleine und grosse Unternehmen. Die Hypothesen werden erneut getrennt für diese beiden Samples überprüft. Wiederum wird die Richtung der Zusammenhänge für sämtliche Hypothesen bestätigt. Die Zusammenhänge sind für grosse Unternehmen durchgängig signifikant, für kleine Unternehmen sind lediglich die Ergebnisse zu *Hypothese H_{2a}* signifikant.⁸ Somit scheinen insbesondere grosse Unternehmen eine systematische Nachhaltigkeitsberichterstattung gemäss der theoretischen Erklärungsansätze vorzunehmen. Dies könnte beispielsweise auf die stärkere Wahrnehmung grosser Unternehmen durch das gesellschaftliche Umfeld und Investoren zurückzuführen sein. Darüber hinaus wird die Berichterstattung für Deutsche und Schweizer Unternehmen getrennt analysiert, da bisherige empirische Studien auf Unterschiede in der Nachhaltigkeitsberichterstattung zwischen verschiedenen Ländern hinweisen (vgl. exemplarisch Hummel 2012, 178). Auch für diese Variante wird die Richtung des Zusammenhangs für sämtliche Hypothesen bestätigt. Die Zusammenhänge sind jedoch durchgängig lediglich für Deutsche und nicht für Schweizer Unternehmen statistisch signifikant.⁹ Eine mögliche Erklärung könnte die LT liefern, wonach das gesellschaftliche Umfeld in der Schweiz gegenüber den Unternehmen weniger kritisch als in Deutschland eingestellt sein könnte. Dies würde mit einem geringeren Legitimitätsdruck auf die Unternehmen in der Schweiz einhergehen.¹⁰

Insgesamt erweist sich der methodische Ansatz zur Messung von Nachhaltigkeitsberichterstattung und Nachhaltigkeitsperformance somit für die Untersuchung der Fragestellungen als geeignet. Auch unsere Argumentation, wonach *sowohl VDT als auch LT* jeweils einen Teil der Nachhaltigkeitsberichterstattung von Unternehmen erklären können, wird durch die Ergebnisse der empirischen Überprüfung sowie die Sensitivitätsanalysen gestützt.

5. Fazit und kritische Diskussion

Ausgangspunkt des vorliegenden Beitrags ist der Zusammenhang zwischen der Nachhaltigkeitsperformance eines Unternehmens und Umfang und Form seiner Nachhaltigkeitsberichterstattung. Mögliche Erklärungsansätze werden im Rahmen der Hypothesenformulie-

⁸ Die Ergebnisse lauten: H_1 : $r=0,23$; $p=0,28$; $n=25$ für kleine Unternehmen bzw. $r=0,45$; $p=0,02$; $n=25$ für grosse Unternehmen. H_{2a} : $r=0,34$; $p=0,09$; $n=25$ für kleine Unternehmen bzw. $r=0,55$; $p=0,00$; $n=25$ für grosse Unternehmen. H_{2b} : $r=-0,18$; $p=0,40$; $n=25$ für kleine Unternehmen bzw. $r=-0,41$; $p=0,04$; $n=25$ für grosse Unternehmen. Alternativ haben wir die Unternehmensgrösse über Umsatz und Marktkapitalisierung (jeweils zum Ende 2010) operationalisiert und erhalten bezüglich Richtung und Signifikanz nahezu identische Ergebnisse.

⁹ Die Ergebnisse lauten: H_1 : $r=0,11$; $p=0,59$; $n=25$ für Schweizer Unternehmen bzw. $r=0,53$; $p=0,01$; $n=25$ für Deutsche Unternehmen. H_{2a} : $r=0,31$; $p=0,14$; $n=25$ für Schweizer Unternehmen bzw. $r=0,55$; $p=0,00$; $n=25$ für Deutsche Unternehmen. H_{2b} : $r=-0,30$; $p=0,15$; $n=25$ für Schweizer Unternehmen bzw. $r=-0,36$; $p=0,08$; $n=25$ für Deutsche Unternehmen.

¹⁰ Die jeweiligen Subsamples auf Basis von Unternehmensgrösse und Landeszugehörigkeit sind heterogen, so dass eine Zuordnung der Effekte zu den beiden Kontrollvariablen plausibel ist. Innerhalb der Gruppe der grossen Unternehmen beträgt der Anteil Schweizer Unternehmen 36% gegenüber einem Anteil von 50% in der Gesamtstichprobe.

rung auf Basis der LT und der VDT abgeleitet. Angesichts methodischer Schwächen bisheriger Studien bei der Operationalisierung der zentralen Untersuchungsvariablen steht im Mittelpunkt des Beitrags die Entwicklung neuer Konzepte zur Messung von Umfang und Form der Nachhaltigkeitsberichterstattung sowie der Nachhaltigkeitsperformance. Erstmalig wird bei der Bewertung der berichteten Informationen durchgängig zwischen quantitativer und verbaler Nachhaltigkeitsberichterstattung unterschieden. Das vorgestellte Messkonzept zur Bestimmung der Nachhaltigkeitsperformance ermöglicht eine objektive und mehrdimensionale Erfassung wie sie nach Kenntnis der Verfasser zum aktuellen Zeitpunkt in dieser Form und in diesem Zusammenhang bislang nicht vorgenommen wurde. Beide Messkonzepte werden hinsichtlich der Erfüllung zentraler Gütekriterien kritisch untersucht und auf eine Stichprobe von 50 Deutschen und Schweizer Grossunternehmen angewandt. Die Ergebnisse zeigen, dass die Messkonzepte für eine Überprüfung der theoretischen Erklärungsansätze geeignet sind. Zudem wird im Unterschied zu bisherigen Untersuchungen deutlich, dass beide Theorien jeweils einen *Teil* der Nachhaltigkeitsberichterstattung erklären können.

Einschränkend muss darauf hingewiesen werden, dass für die statistischen Auswertungen aufgrund der relativ geringen Stichprobengrösse der Einsatz multivariater Analysemethoden nicht zweckmäßig ist. Hier bieten sich Anhaltspunkte für weitere Forschungsarbeiten. Neben einer Vergrösserung der Datenbasis ist insbesondere die Berücksichtigung einer Reihe von Kontrollvariablen für eine abschliessende Analyse der Zusammenhänge erforderlich. So weisen die Ergebnisse der vorliegenden Untersuchung auf grundlegende Unterschiede zwischen Deutschen und Schweizer Unternehmen hin. Durch eine Ausweitung der Datenbasis auf weitere Länder könnte der Einfluss länderspezifischer Besonderheiten im Rahmen zukünftiger Untersuchungen näher betrachtet werden. Zudem könnte die Erfassung der Nachhaltigkeitsperformance weiter ausdifferenziert werden. Denkbar ist beispielsweise eine branchenspezifischere Beurteilung anstelle der einfachen Unterscheidung in Industrie- und Dienstleistungsunternehmen.

Letztlich beruhen die Daten zur Messung der Nachhaltigkeitsperformance auf Angaben der berichtenden Unternehmen. Die Ausführungen in diesem Beitrag zeigen, dass auch bei Anwendung der GRI-Richtlinien zur Nachhaltigkeitsberichterstattung eine Vergleichbarkeit der berichteten Daten nicht ohne Weiteres gegeben ist. Dadurch wird die Auswahl der Indikatoren für das Messkonzept stark limitiert. Hier könnten einheitliche und bindende Vorschriften zum Inhalt und zur Prüfung der Nachhaltigkeitsberichterstattung deren Transparenz und Qualität deutlich erhöhen. So würde die Erfassung der Nachhaltigkeitsperformance und ein darauf basierender Vergleich verschiedener Unternehmen einfacher.

Literaturhinweise

- Al-Tuwaijri, S.A./Christensen, T.E./Hughes, K.E. (2004): The relations among environmental disclosure, environmental performance, and economic performance: a simultaneous equations approach, in: Accounting, Organizations and Society, Jg. 29, Nr. 5, S. 447–471.
- Ashforth, B.E./Gibbs, B.W. (1990): The double edge of legitimization, in: Organization Science, Jg. 1, Nr. 2, S. 177–194.
- Bewley, K./Li, Y. (2000): Disclosure of environmental information by Canadian manufacturing companies: A voluntary disclosure perspective, in: Advances in Environmental Accounting & Management, Jg. 1, S. 201–226.

- Bortz, J./Döring, N. (2006): Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler, 4. Auflage, Heidelberg.
- Breuer, W./Breuer, C./Heldt, C./Lackes, R./Siepermann, M. (2010): Performance, in: Gabler Wirtschaftslexikon, 17. Auflage, Wiesbaden, S. 2328.
- Chatterji, A.K./Levine, D.I./Toffel, M.W. (2009): How Well Do Social Ratings Actually Measure Corporate Social Responsibility?, in: Journal of Economics & Management Strategy, Jg. 18, Nr. 1, S. 125–169.
- Cho, C.H./Guidry, R.P./Hageman, A.M./Patten, D.M. (2012): Do actions speak louder than words? An empirical investigation of corporate environmental reputation, in: Accounting, Organizations and Society, Jg. 37, Nr. 1, S. 14–25.
- Cho, C.H./Patten, D.M. (2007): The role of environmental disclosures as tools of legitimacy: A research note, in: Accounting, Organizations and Society, Jg. 32, Nr. 7, S. 639–647.
- Cho, C.H./Patten, D.M./Roberts, R.W. (2006): Corporate Political Strategy: An Examination of the Relation between Political Expenditures, Environmental Performance, and Environmental Disclosure, in: Journal of Business Ethics, Jg. 67, Nr. 2, S. 139–154.
- Clarkson, P.M./Li, Y./Richardson, G.D./Vasvari, F.P. (2008): Revisiting the relation between environmental performance and environmental disclosure: An empirical analysis, in: Accounting, Organizations and Society, Jg. 33, Nr. 4, S. 303–327.
- Clarkson, P.M./Overell, M.B./Chapple, L. (2011): Environmental Reporting and its Relation to Corporate Environmental Performance, in: Abacus, Jg. 47, Nr. 1, S. 27–60.
- Davis, K. (1973): The case for and against business assumption of social responsibilities, in: Academy of Management Journal, Jg. 16, Nr. 2, S. 312–322.
- Dawkins, C./Fraas, J.W. (2011): Coming Clean: The Impact of Environmental Performance and Visibility on Corporate Climate Change Disclosure, in: Journal of Business Ethics, Jg. 100, Nr. 2, S. 303–322.
- de Villiers, C./van Staden, C.J. (2006): Can less environmental disclosure have a legitimizing effect? Evidence from Africa, in: Accounting, Organizations and Society, Jg. 31, Nr. 8, S. 763–781.
- Deegan, C. (2002): Introduction: The legitimising effect of social and environmental disclosures – a theoretical foundation, in: Accounting, Auditing & Accountability Journal, Jg. 15, Nr. 3, S. 282–311.
- Dowling, J./Pfeffer, J. (1975): Organisational legitimacy: social values and organisational behavior, in: Pacific Sociological Review, Jg. 18, Nr. 1, S. 122–136.
- Elkington, J. (1999): Cannibals with Forks: The Triple Bottom Line of 21st Century Business, Oxford.
- Fekrat, M.A./Inclan, C./Petroni, D. (1996): Corporate Environmental Disclosures: Competitive Disclosure Hypothesis Using 1991 Annual Report Data, in: The International Journal of Accounting, Jg. 31, Nr. 2, S. 175–195.
- Global Reporting Initiative (2012): What is GRI?, unter <https://www.globalreporting.org/information/about-gri/what-is-GRI/Pages/default.aspx> (Abfragedatum 1.11.2012).
- Gray, R./Kouhy, R./Lavers, S. (1995): Corporate social and environmental reporting: A review of the literature and a longitudinal study of UK disclosure, in: Accounting, Auditing & Accountability Journal, Jg. 8, Nr. 2, S. 47–77.
- Healy, P.M./Palepu, K.G. (2001): Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature, in: Journal of Accounting and Economics, Jg. 31, Nr. 1-3, S. 405–440.

- Hoffmann, T. (2011): Unternehmerische Nachhaltigkeitsberichterstattung. Eine Analyse des GRI G3.1-Berichtsrahmens, Lohmar.
- Hughes, S.B./Anderson, A./Golden, S. (2001): Corporate environmental disclosures: are they useful in determining environmental performance?, in: Journal of Accounting and Public Policy, Jg. 20, Nr. 3, S. 217–240.
- Hummel, K. (2012): Internationaler Vergleich der Nachhaltigkeitsberichterstattung, in: Controlling, Jg. 24, Nr. 3, S. 175–180.
- Ingram, R.W./Frazier, K.B. (1980): Environmental Performance and Corporate Disclosure, in: Journal of Accounting Research, Jg. 18, Nr. 2, S. 614–622.
- Kolk, A. (2010): Determinants of the Adoption of Sustainability Assurance Statements: An International Investigation, in: Business Strategy and the Environment, Jg. 19, Nr. 3, S. 182–198.
- Li, Y./Richardson, G.D./Thornton, D. (1997): Corporate Disclosure of Environmental Liability Information: Theory and Evidence, in: Contemporary Accounting Research, Jg. 14, Nr. 3, S. 435–474.
- Milne, M.J./Patten, D.M. (2002): Securing organizational legitimacy: An experimental decision case examining the impact of environmental disclosures, in: Accounting, Auditing & Accountability Journal, Jg. 15, Nr. 3, S. 372–405.
- Morhardt, J.E./Baird, S./Freeman, K. (2002): Scoring Corporate Environmental and Sustainability Reports Using GRI 2000, ISO 14031 and other Criteria, in: Corporate Social Responsibility and Environmental Management, Jg. 9, Nr. 4, S. 215–233.
- O'Donovan, G. (2002): Environmental disclosures in the annual report: Extending the applicability and predictive power of legitimacy theory, in: Accounting, Auditing & Accountability Journal, Jg. 15, Nr. 3, S. 344–371.
- Patten, D.M. (2002): The relation between environmental performance and environmental disclosure: a research note, in: Accounting, Organizations and Society, Jg. 27, Nr. 8, S. 763–773.
- Quick, R./Knocinski, M. (2006): Nachhaltigkeitsberichterstattung, in: Zeitschrift für Betriebswirtschaft, Jg. 76, Nr. 6, S. 615–650.
- Rockness, J.W. (1985): An Assessment of the Relationship between US Corporate Environmental Performance and Disclosure, in: Journal of Business Finance & Accounting, Jg. 12, Nr. 3, S. 339–354.
- Sethi, S.P. (1978): Advocacy Advertising – The American Experience, in: California Management Review, Jg. 21, Nr. 1, S. 55–68.
- Suchman, M.C. (1995): Managing Legitimacy: Strategic and Institutional Approaches, in: The Academy of Management Review, Jg. 20, Nr. 3, S. 571–610.
- van der Laan Smith, J./Adhikari, A./Tondkar, R.H. (2005): Exploring differences in social disclosures internationally: A stakeholder perspective, in: Journal of Accounting and Public Policy, Jg. 24, Nr. 2, S. 123–151.
- Verrecchia, R.E. (1983): Discretionary Disclosure, in: Journal of Accounting and Economics, Jg. 5, Nr. 1, S. 179–194.
- Wiseman, J. (1982): An Evaluation of Environmental Disclosures Made in Corporate Annual Reports, in: Accounting, Organizations and Society, Jg. 7, Nr. 1, S. 53–63.

Katrin Hummel, Dr. rer. pol., ist Oberassistentin am Lehrstuhl für Accounting, insbesondere Unternehmensrechnung und Controlling am Institut für Betriebswirtschaftslehre der Universität Zürich.

Christian Schlick, Dipl.-Kfm., ist Assistent am Lehrstuhl für Accounting, insbesondere Unternehmensrechnung und Controlling am Institut für Betriebswirtschaftslehre der Universität Zürich.

Anschrift: Universität Zürich, Institut für Betriebswirtschaftslehre, Lehrstuhl für Accounting, insbes. Unternehmensrechnung und Controlling, Plattenstr. 14, CH-8032 Zürich, Tel.: +41 (0) 44/634-29-83, Fax: +41 (0) 44/634-49-12, E-Mails: katrin.hummel@business.uzh.ch; christian.schlick@business.uzh.ch