



JUNIOR MANAGEMENT SCIENCE

Advisory Editorial Board:

DOMINIK VAN AAKEN
FREDERIK AHLEMANN
CHRISTOPH BODE
ROLF BRÜHL
JOACHIM BÜSCHKEN
LEONHARD DOBUSCH
RALF ELSAS
DAVID FLORYSIK
GUNTHER FRIEDL
WOLFGANG GÜTTEL
CHRISTIAN HOFMANN
KATJA HUTTER
LUTZ JOHANNING
STEPHAN KAISER
ALFRED KIESER
NATALIA KLIEWER
DODO ZU KNYPHAUSEN-AUFSEß
SABINE T. KÖSZEGI
ARJAN KOZICA
TOBIAS KRETSCHMER
HANS-ULRICH KÜPPER
REINER LEIDL
ANTON MEYER
MICHAEL MEYER
GORDON MÜLLER-SEITZ
BURKHARD PEDELL
MARCEL PROKOPCZUK
TANJA RABL
SASCHA RAITHEL
ASTRID REICHEL
KATJA ROST
MARKO SARSTEDT
DEBORAH SCHANZ
ANDREAS G. SCHERER
STEFAN SCHMID
UTE SCHMIEL
CHRISTIAN SCHMITZ
PHILIPP SCHRECK
GEORG SCHREYÖGG
LARS SCHWEIZER
DAVID SEIDL
THORSTEN SELLHORN
ANDREAS SUCHANEK
ORESTIS TERZIDIS
ANJA TUSCHKE
SABINE URNIK
STEPHAN WAGNER
BARBARA E. WEIßENBERGER
ISABELL M. WELPE
HANNES WINNER
CLAUDIA B. WÖHLE
THOMAS WRONA
THOMAS ZWICK

Mischa Haberthür , What makes a responsible leader? – An empirical analysis of how personality characteristics affect leadership behaviour	1
Lea Katharina Haffke , Zur experimentellen Analyse des Zusammenhangs von Feedback, Selbstwirksamkeit und Kreativität	38
Iris Krampe , Besteuerung unternehmensverbundener Stiftungen	55
Lukas Erhard , The Effects of Battery Storage on Risk and Cost of Capital of Wind Park Investments	74
Bernhard Guetz , Konzeptualisierung und Operationalisierung von Patientenzufriedenheit mit Hausärztinnen und Hausärzten	106
Sebastian Schär , State-of-the-Art dynamischer Methoden zur multikriteriellen Entscheidungsunterstützung	145



What makes a responsible leader? - An empirical analysis of how personality characteristics affect leadership behaviour

Mischa Haberthür

Universität Zürich

Abstract

This master's thesis contributes to the ongoing discussion on responsible leadership and its antecedents by empirically analysing the relationship between various traits and responsible leadership effectiveness. In a quantitative experimental research study, the participants (62 men and 32 women) first filled out an online survey assessing their gender, age, and major field of study as well as their score in the personality dimensions honesty-humility, openness to experience, empathic concern, perspective taking, and holism. Afterwards, their second-to-fourth digit ratio was measured, and men had the option of providing a saliva sample which was used to determine their testosterone level. Finally, the participants' performance in the leadership roles professional, facilitator, citizen, idea-provider, and total was assessed with a think-aloud protocol. By performing a multiple linear regression analysis, I found that empathic concern positively influences the roles facilitator and total, holism the role citizen, and perspective taking the role idea-provider. I also found that men and psychology students are less likely to perform well in the role idea-provider and that openness to experience negatively affects facilitator.

Keywords: Responsible leadership, Ethical leadership, Corporate social responsibility, Stakeholder approach, Personality characteristics

1. Introduction

Over the course of the past few decades, political, economic, social, environmental, and technological activities have become increasingly interrelated on a worldwide scale (Figge and Martens (2014)). This ongoing process of globalisation has had a tremendous impact on the size and structural composition of companies as well. Nowadays, it is common for markets to be dominated by one or several multinational firms, which offer their products in a large variety of countries. Not only do these firms hold considerable market power, their influence frequently transcends a purely economic sphere. Many of the largest global corporations operate in countries with underdeveloped regulatory frameworks, granting them a lot of leeway in their regular business activities. Furthermore, their world-spanning reach impedes the nation states' capability of moderating their potentially harmful behaviour (Voegtlin et al. (2012)). This development in combination with the emergence of corporate scandals, such as the Enron case or the BP oil spill, has sparked a public debate on the role a market organisation should occupy in society and the responsibilities it has towards its environment.

The public discourse on corporate responsibility has been extended to incorporate the individual responsibility of leadership figures as well. Various researchers have argued that the traditional understanding of leadership effectiveness, which tends to put the focus on creating shareholder value and maximising profits, is inadequate for a globalised and interwoven business environment, among them Maak and Pless (2006), Patzer (2009), Waldman and Galvin (2008), and Waldman and Siegel (2008). Drawing on research from the fields of stakeholder theory and corporate social responsibility, these authors posit that previous leadership theories have neglected the importance of ethical behaviour and responsibility. They argue that the challenges presented by the globalised business world call for a new paradigm in leadership research, a concept that has been labelled responsible leadership.

While responsible leadership does not yet have a unifying definition that is accepted by all researchers in the field, it is generally considered to be related to fulfilling stakeholder needs and expectations. The underlying argument to this approach is that a leader's responsibility does not begin and end solely with the shareholders. Instead, he or she has an obli-

gation towards any person or group which is affected by the company's activities. To be truly responsible, a leader must therefore be willing to serve all stakeholders to the best of his or her abilities (Maak (2007)).

Based on these theoretical arguments, Voegtlin and Schwab (2015) have developed a multi-dimensional model of responsible leadership, in which leaders take on one of three different roles depending on which stakeholder group they engage. A responsible leader must be able to act as a professional towards shareholders or other owners of the company, as a facilitator towards colleagues and subordinates, and as a citizen towards external stakeholders. The better a leader is able to perform in all three roles simultaneously, the more he or she is considered to act responsibly.

While the link between responsibility and a focus on stakeholder demands is generally accepted in the literature, there is some dispute over what exactly motivates leaders to act in a responsible manner (Miska et al. (2014)). Some researchers suggest that responsible leadership is simply the result of calculated profit maximisation. Leaders engage stakeholders in order to gain a competitive advantage and increase the value of their business (Porter and Kramer (2006)). Other researchers posit that responsible leadership is at least partly informed by a leader's personal values or sense of duty (Miska et al. (2014); Waldman and Galvin (2008)).

The second approach would indicate that responsible leadership behaviour is not simply a rational choice that all leaders calculate in an equal manner, but is to some extent dependent on a leader's individual personality. This suggests that specific personality characteristics can increase or decrease the likelihood of a leader acting responsibly. I find some support for this assertion in the theoretical literature (Freeman and Auster (2011); Miska et al. (2014); Pless and Maak (2011)), but there is as of yet no clear consensus on which personality characteristics influence responsible leadership. The aim of my master's thesis is to help fill this gap in the current research by performing a quantitative empirical analysis to determine whether there are traits which have an impact on responsible leadership effectiveness.

To achieve this goal, I first established a set of ten traits which I argue could potentially influence responsible leadership behaviour. These are a person's gender, age, whether or not he or she is a psychology student, his or her testosterone level and second-to-fourth digit ratio (2D:4D) as well as his or her score in the personality dimensions honesty-humility, openness to experience, empathic concern, perspective taking, and holism. Additionally, I defined a set of outcome variables that can be used to measure responsible leadership effectiveness directly. For this, I relied on the multi-dimensional leadership model by Voegtlin and Schwab (2015) and its three roles professional, facilitator, and citizen as well as the two additional roles total and idea-provider. Using statistical regression analysis, I constructed a variety of multiple linear regression models that can be used to show whether any of the observed personality characteristics influence responsible leadership behaviour.

The data for these models was collected in collaboration with members of the Chair of Clinical Psychology and Psychotherapy and the Chair of Foundations of Business Administration and Theories of the Firm at the University of Zurich (UZH). We performed a scenario-based research study, in which 94 participants completed a survey on their individual personality traits and were then graded by us according to the responsible leadership effectiveness they displayed in a think-aloud protocol.

This master's thesis is divided into five chapters. In this chapter, the Introduction, I have given a brief overview of the thesis' subject, discussed my basic premise, and introduced the research goal I intend to achieve. Chapter 2 will focus on the theoretical foundations of my empirical regression study. In Subchapter 2.1, I will offer a general definition of leadership and discuss various research methods that have been used to measure it. I will then introduce the comparatively new field of responsible leadership, including the multi-dimensional model by Voegtlin and Schwab (2015), and highlight its differences from traditional approaches. Lastly, I will use these theoretical concepts to derive my research question. Subchapter 2.2 discusses the individual traits of my empirical regression analysis in greater detail. I will introduce each trait separately and form a hypothesis about its potential impact on responsible leadership effectiveness based on previous research.

Chapter 3 is concerned with the specific methods I used for collecting and analysing my data. I will talk about the research study conducted by members of the Chair of Clinical Psychology and Psychotherapy, the Chair of Foundations of Business Administration and Theories of the Firm, and myself, explaining our study design and the reasoning behind it. Subchapter 3.1 talks about the selection of participants and our recruitment process. Subchapter 3.2 is focused on the different variables we measured, while Subchapter 3.3 gives a brief overview of the actual procedures of the study. Finally, Subchapter 3.4 talks about the delimitations and limitations of our research study and their potential implications.

In Chapter 4, I will present the results of my empirical regression analysis. Each subchapter focuses on a different role of responsible leadership; Subchapter 4.1 talks about the models for the role of professional, Subchapter 4.2 about facilitator, Subchapter 4.3 about citizen, Subchapter 4.4 about idea-provider, and Subchapter 4.5 about total. Additionally, Subchapter 4.6 analyses the effects of testosterone levels on all five roles of responsible leadership.

Finally, Chapter 5 is going to feature the discussion and interpretation of my results from Chapter 4. I will once again look at each trait separately, compare the outcomes of my multiple linear regression models with the expected outcomes I have formulated in my original hypotheses, and determine whether there exist any congruencies or divergences. Subchapter 5.1 discusses the demographic and biological traits, while Subchapter 5.2 focuses on the personality characteristics. In Subchapter 5.3, I will analyse more general observations about my regression models that are not directly related to a single trait. Lastly, Subchapter 5.4

presents the conclusions of my thesis. I will give a brief summary of the most notable findings and highlight potential avenues that future research could focus on.

2. Theoretical foundations

This chapter discusses the theoretical frameworks which form the basis of my empirical research study. It is divided into two parts. In Subchapter 2.1, I will discuss the current state of the research on leadership and derive my generalised research question from it. In Subchapter 2.2, I will turn this research question into a specific set of hypotheses and introduce the traits I intend to measure.

2.1. Leadership

This subchapter discusses the current research in the field of leadership. In Part 2.1.1, I will give a brief overview of the traditional approaches to measuring leadership effectiveness. Part 2.1.2 will introduce the concept of responsible leadership, offer a comprehensive definition of the term, and discuss its implications. Part 2.1.3 concerns itself with the multi-dimensional leadership model and its applications to responsible leadership research. Finally, in Part 2.1.4, I will present my generalised research question and show how it was derived from combining the traditional approaches with the comparatively new field of responsible leadership.

2.1.1. Traditional approaches

The concept of leadership and its implications for management practices and organisational structures has always been one of the most extensively studied topics in the social sciences (Day and Antonakis (2012)). While the term leadership has different connotations depending on the context it is used in, researchers generally see it as a process of influencing other people in order to achieve a common goal (Yukl (2002)). A leader therefore is an individual capable of rallying followers and directing their efforts towards a specific task (Winston and Patterson (2006)).

The precise nature of such leaders and the question of what separates them from their followers has been the subject of extensive debates in the field. One of the earliest attempts to explain this phenomenon was the Great Man theory developed by the nineteenth century historian Thomas Carlyle, which posits that effective leadership is an inherent quality of specific individuals (Judge et al. (2002)). In this approach, leaders and followers are differentiated solely by certain heritable traits. Thus, great leaders are born rather than made (Hoffman et al. (2011)).

While the notion that only selected individuals are capable of becoming great leaders gradually fell out of favour over the course of the twentieth century, the idea that effective leadership results from specific traits has remained popular (Zaccaro (2007)). The trait perspective became the dominant approach in leadership research during the first half of the twentieth century, with researchers seeking to explore the connection between an individual's personality and his or her

effectiveness as a leader (Jago (1982)). Numerous studies attempted to quantify a set of traits that positively correlate with leadership effectiveness, operating under the assumption that identifying those traits could assist in selecting already qualified leaders or in fostering leadership behaviour within individuals who exhibit potential.

However, the empirical results from these studies often proved to be less significant than expected (Judge et al. (2002)). Many of the proposed traits did not show any correlations with leadership effectiveness, and for those that did, the effect often could not be replicated in subsequent studies. A seminal literature review by Stogdill (1948) found that while some traits indeed appear to impact leadership effectiveness, their importance is overstated when compared to more situational factors. The issue was further exacerbated by the fact that researchers sometimes used differing definitions for the same traits, which made comparing studies a difficult task. Additionally, there was no clear consensus on what exactly constitutes effective leadership and how it should be measured (Derue et al. (2011)).

The apparent problems of the purely trait-focused perspective caused a realignment in the field and led to the development of entirely new approaches to understanding leadership. Among them were behavioural theories, which claimed that effective leadership is the result of a leader's actions rather than his or her inherent qualities (Yukl (1989)). Proponents of this approach argued that there exist specific behaviours which generate effective leadership (House and Aditya (1997)). By analysing how leaders interact with their followers and replicating their successful strategies, leadership behaviour could therefore be learned by anyone.

As a reaction to both trait and behavioural perspectives on leadership, various contingency theories emerged in the second half of the twentieth century, for instance path-goal theory, situational leadership theory, and cognitive resource theory (House and Aditya (1997)). While these theories differed from each other in certain aspects, they all agreed that previous attempts at explaining effective leadership were too simplistic because they failed to take contextual factors into consideration (Judge et al. (2002)). Adherents of contingency theories argued that there are no universal behaviours or personality characteristics which always result in effective leadership. Instead, situational aspects, such as the social environment or a leader's degree of control, determine whether any given trait or action will contribute to the success of an organisation (Vroom and Jago (2007)).

Over the past few decades, trait theory has experienced a surge in popularity. By statistically analysing earlier studies, researchers found that the link between individual traits and leadership effectiveness was often more pronounced than had previously been assumed (Zaccaro et al. (2004)). In addition, the emergence of charismatic and transformative leadership once again highlighted the importance of individual characteristics on the leadership process (Zaccaro (2007)). More recently, several empirical studies have discovered positive correlations between specific personality attributes and leadership effectiveness, for instance Judge

et al. (2002), Judge et al. (2004), and McCormack and Mellor (2002).

What separates these new studies from traditional trait theory is that they no longer consider traits to be solely responsible for leadership effectiveness. Instead, they acknowledge the importance of situational effects and behavioural inputs, arguing that a comprehensive assessment of leadership needs to integrate all relevant factors (Derue et al. (2011)). The current research also puts a greater emphasis on harmonising the measurements used in empirical studies. Early research in the field was heavily fragmented, as studies often looked at completely different traits or analysed similar traits under different names (Judge et al. (2002)). More recent approaches try to mitigate this problem by focusing on measurements that are already well-established in psychological research, for instance the five-factor model (FFM) of personality (Judge et al. (2002)).

2.1.2. Responsible leadership

As has been shown in the previous part, leadership is generally considered to be a means of fulfilling a particular task or achieving a specific goal. However, the exact nature of this goal is not always clearly defined and strongly depends on the individual context. For instance, a military leader's objective might be to win a particular battle or ensure that casualties among his or her troops are minimised, whereas a civil rights leader would want to raise public awareness for his or her cause and implement social reforms. In both examples, effective leadership is important for reaching the desired goal, but a leadership approach which is successful in one situation may not necessarily produce the same results in the other.

Within the context of organisations competing in a market environment, a lot of empirical research has focused on leaders in the lower management, whose goals are predetermined by their superiors (House and Aditya (1997)). In those cases, leadership effectiveness has often been measured by looking at how well the managers accomplished the objectives set for them, while sometimes also evaluating the satisfaction of their employees. However, this approach is of limited use when analysing the leadership effectiveness of CEOs, who are not only able, but also expected to set the strategic goals for the entire organisation (Finkelstein and Boyd (1998)).

When evaluating the leadership effectiveness of a CEO, both researchers and the general public often tend to adopt a shareholder value approach (Waldman and Galvin (2008)). In this view, a CEO's sole responsibility is towards the shareholders of his or her company. As their direct employee, the CEO must ensure that the demands of the shareholders are met, a notion that is generally equated with maximising profits, stock prices, and future growth potential (Carson (1993)). Other considerations, such as employee or customer satisfaction, are only important in so far as that they contribute to maximising the shareholder value. Proponents of the shareholder value approach oftentimes argue that the focus on pure profit maximisation is not only beneficial to

the organisation itself, but also to society at large (Waldman and Galvin (2008)). Nevertheless, the theory has been heavily criticised for neglecting various groups who are of central importance to an organisation's continued operations (Russo and Perrini (2010)). Several researchers have pointed out that a pure shareholder-orientation does not adequately reflect actual business operations and that CEOs should instead try to incorporate other stakeholders in their decision-making process (Laplume et al. (2008)).

Criticism of the shareholder approach has not been solely confined to an academic setting. Corporate scandals, such as the Enron accounting fraud in 2001 or the BP oil spill in the Gulf of Mexico in 2010, have caused the public to take greater notice of corporate leaders, thereby sparking discussions about the specific role a business should occupy in society (Maak (2007)). The resulting public backlash led to a considerable loss of corporate legitimacy as well as a general distrust of leadership figures (Voegtlin et al. (2012)).

The notion of CEOs as pure profit-optimisers is further challenged by the ongoing globalisation process and the resulting lack of clearly-defined governance structures (Voegtlin et al. (2012)). Since businesses nowadays are often operating in areas with underdeveloped regulatory frameworks, the public has come to expect that they not only fulfil their basic obligations towards their shareholders, but also engage in behaviour which does not fall under the purview of standard market activities, such as combating corruption or poverty (Pless and Maak (2011)). Multinational firms thus transform from mere market participants into political actors (Voegtlin et al. (2012)).

With globalisation and scandals pushing the traditional understanding of corporate leadership to its limits, researchers have been trying to expand the term to incorporate these new challenges. Inspired by the academic discussions on stakeholder theory and corporate social responsibility, a new trend has emerged in the leadership literature in recent years; that of responsible leadership (Pless (2007)). As Waldman and Galvin (2008) note in their analysis of previous leadership theories:

We also propose that the responsibility element is missing from these descriptors, and that it is actually this element that is at the heart of what effective leadership is all about. In a nutshell, to not be responsible is to not be effective as a leader. (p. 327)

Offering a comprehensive definition of responsible leadership is not an easy task, as different authors approach the term from their own unique perspectives (Miska et al. (2014)). In the following paragraphs, I will attempt to summarise the current state of the field by introducing two specific definitions, both of which either have been directly cited or whose meaning is reflected in multiple other research papers. I will then compare and contrast the two definitions to gain a greater understanding of what the responsible leadership concept actually entails. The first definition comes from an article by Maak and Pless (2006):

Responsible leadership is a relational and ethical phenomenon, which occurs in social processes of interaction with those who affect or are affected by leadership and have a stake in the purpose and vision of the leadership relationship. (p. 103)

The second definition is from an article by Waldman (2011):

[Responsible leadership] considers what managers should do in an effort to take into account the needs of stakeholders, other than shareholders, who may have legitimate interests in a firm's activities. (p. 77)

What these definitions have in common is that they embed leadership in the context of stakeholder interactions. The authors argue that a corporate leader has a responsibility not only to the company's shareholders, but to every person or group that is affected by the company's business activities. A responsible leader is therefore one who is able to fulfil stakeholder needs and expectations (Maak (2007)).

Both definitions also agree that responsible leadership is at least to some extent a question of morality. Maak and Pless (2006) explicitly refer to it as an ethical phenomenon, while Waldman (2011) draws comparisons to the related field of ethical leadership in his own article. This implies that responsible leadership transcends a purely self-serving perspective on business activities. A leader should not establish good relationships with stakeholders merely so they will grant legitimacy; rather, legitimacy is the natural result of a leader being a responsible member of the community. What separates responsible leadership from the concept of ethical leadership is that the former does not focus on a leader's own ethical norms and values. Instead, an ethical solution to a problem is created through the interaction between leaders and internal as well as external stakeholders (Waldman and Galvin (2008)).

One minor aspect in which the two presented definitions somewhat differ from each other is their conceptualisation of the actual leadership process. Maak and Pless (2006) explicitly stress the importance of direct interactions with stakeholders as a means of becoming a responsible leader. Their view is informed by the concept of deliberative democracy, which states that corporate legitimacy is gained through approaching stakeholders as equals and engaging them in a constructive dialogue (Palazzo and Scherer (2006)). In the ideal-type scenario, finding an appropriate solution to an issue would be done through a democratic process in which every affected person or entity has a voice. Waldman (2011), on the other hand, only states that a manager should consider the needs of stakeholders, but leaves open how strongly said stakeholders should be integrated into the final decision-making process. Under his definition, a leader could theoretically decide in an authoritative manner and still be considered responsible, as long as he gave due consideration to the stakeholders' demands in his deliberations.

2.1.3. Multi-dimensional leadership model

Despite their semantic differences, the definitions of Maak and Pless (2006) and Waldman (2011) seem to roughly agree on what responsible leadership is and how it can be considered distinct from other leadership concepts. Their basic assumption that the increasingly complex business environment necessitates a reconceptualisation of corporate leadership with a greater emphasis on stakeholder demands is reflected in many other writings on the subject, for instance Cameron (2011), Patzer (2009), Voegtlin (2011), and Waldman and Siegel (2008).

However, there is as of yet no clear consensus on how these theoretical concepts can be optimally implemented into practice, especially considering that stakeholders often have different and seemingly conflicting demands (Voegtlin and Schwab (2015)). A manager who attempts to incorporate all affected parties in the decision-making process may soon realise that he or she is faced with significant trade-offs. For instance, consumers usually want to pay as little as possible for a given product, but this desire may conflict with the employees demanding fair wages for their labour. Alternatively, the public's wish for stronger environmental protection can be difficult to reconcile with the shareholders' demands for continually increasing profits and growth. In addition to these inherent contradictions, managers are often faced with resource constraints as well. Temporal and financial limitations may make it difficult to give every stakeholder group the attention it requires or deserves.

Some researchers have proposed that these issues can be resolved by prioritising stakeholders based on their salience (Bundy et al. (2013)). A famous example of this approach is the framework developed by Mitchell et al. (1997), which measures stakeholder salience on the basis of three criteria: power, legitimacy, and urgency. They argue that stakeholders who possess all three attributes should be given high priority, while those with only one attribute are of lesser importance. This approach allows leaders to resolve conflicting stakeholder demands by focusing on whichever group matters more to the firm's operations. However, stakeholder salience is not always easy to determine and constantly fluctuating, which can make it difficult to objectively assess. Additionally, salience prioritisation does not seem to conform to the normative literature on responsible leadership, which tends to argue that all stakeholder demands should be recognized, regardless of their importance (Maak and Pless (2006); Waldman and Galvin (2008)).

In an attempt to approach the issue from a new perspective, Voegtlin and Schwab (2015) argue that the apparent gap between theoretical literature and practical considerations is not necessarily insurmountable. They propose that leaders might be able to deal with complex and paradoxical situations by becoming more complex themselves. Through the use of innovative approaches, a leader could reconcile contradictory stakeholder demands, thereby turning them from a seemingly unsolvable problem into a win-win situation for all involved parties. Thus, responsible leadership re-

quires a new paradigm of management which places a strong emphasis on behavioural complexity (Voegtlin and Schwab (2015)).

Building on these arguments and drawing inspiration from psychological contracts theory as well as stewardship theory, Voegtlin and Schwab (2015) developed a multi-dimensional model of leadership. They suggest that a leader has different responsibilities towards different groups. As such, managers in a globalised, highly complex business environment cannot afford to be inflexible; instead, they need to be able to take on different roles depending on which stakeholders they are addressing. With regards to shareholders or other owners of the company, a responsible leader must be able to act as a professional, fulfilling his or her fiduciary duty and ensuring that tasks are performed effectively and efficiently. Towards colleagues and subordinates, a responsible leader takes on the role of facilitator, caring for the other members of the workplace and creating a fair work environment. Finally, a responsible leader must act as a citizen towards external stakeholders, taking sustainability into consideration and analysing the firm's long-term effects on society (Voegtlin and Schwab (2015)). These three roles are summarised in Table 1 below.

Voegtlin and Schwab (2015) infer a causal relationship between their model and effective responsible leadership. They argue that the main challenge managers face when attempting to become responsible leaders is finding a balance between the roles of professional, facilitator, and citizen, thereby serving all stakeholder groups to the best of their abilities. The better a leader is able to perform multiple roles simultaneously, by observing high behavioural complexity and finding innovative solutions that satisfy all stakeholders, the more he or she is considered to act responsibly. Thus, their model does not only serve as a potential guideline for managers who wish to incorporate responsible leadership considerations into their business activities, it also offers a new perspective on how to empirically assess responsible leadership effectiveness.

2.1.4. Research question

In Part 2.1.1, I introduced the trait theory of leadership with its basic assumption that specific traits of individuals serve as antecedents to effective leadership. Empirical research in the field has focused on a large variety of personality characteristics that could potentially influence leadership behaviour, such as self-confidence, aggressiveness, dominance, and integrity (Judge et al. (2002)). So far, trait theory has predominantly worked with a traditional understanding of leadership, in which effectiveness is measured by how well the manager in question accomplished his or her given tasks or through surveying employee satisfaction (House and Aditya (1997)). However, proponents of responsible leadership posit that this understanding is too narrow for a globalised, highly interwoven business environment, arguing that it should be expanded to include aspects which are not reflected in a purely strategic approach. It seems reasonable to assume that this reconceptualisation would also result in a

change of the antecedents. In other words, the traits which help one become an effective leader under a shareholder value paradigm are not necessarily the same traits that aid in becoming an effective responsible leader.

The question of what makes a responsible leader has been the subject of extensive debates in the theoretical literature. Various personality traits have been proposed as potential antecedents for effective responsible leadership, including relational and ethical intelligence (Maak and Pless (2006)), certain moral predispositions (Voegtlin et al. (2012)), authenticity (Freeman and Auster (2011)), virtuousness (Cameron (2011)), and cognitive complexity (Maak et al. (2016)). However, there is as of yet little empirical research into how an individual's personality affects responsible leadership effectiveness (Fernando (2016)). This may be partly due to the complexity in measurement and data collection, as traditional follower surveys are not necessarily comprehensive enough to assess the full spectrum of responsible leadership activities (Waldman (2011)).

The goal of my master's thesis is to help fill this gap in the current empirical research and contribute to the ongoing academic discussion through performing a quantitative empirical analysis. By drawing on the well-established trait theory and applying its methods to the comparatively new field of responsible leadership, I aim to answer the following research question:

Research question: Which personality characteristics increase the likelihood that a leader will act responsibly?

As the field of responsible leadership is still in its early, formative years (Fernando (2016)), finding an answer to this question could provide valuable information on which directions are worth pursuing in future research; either by corroborating already established theories or by highlighting new avenues that hitherto have not been considered in the literature. Furthermore, exploring the link between personality characteristics and responsible leadership behaviour may offer practical advice for firms and their boards of directors. Finding the right person for a management position has always been a challenging task (Beck and Harter (2014)), and the increasing complexity of the globalised business environment compounds this issue even further. The consequences of a bad selection can be disastrous, as evidenced by corporate scandals, such as the Enron case, and the resulting loss of corporate legitimacy (Voegtlin et al. (2012)). Finding a relationship between certain traits and leadership effectiveness could help mitigate this problem, as it would give firms a better understanding of which qualities they should be looking for in an aspiring manager or promoting in their own organisational culture.

To answer my research question, I will proceed in two steps. Firstly, I define a group of specific traits that can be expected to have an influence on responsible leadership behaviour. My choice of traits and the justification for their selection will be discussed in-depth in the following subchapter. Secondly, I establish a set of criteria that allow

Table 1: Responsible leadership roles (Voegtlin and Schwab (2015))

Responsible leadership role	Moral obligation (that develops out of:)	Accountability toward stakeholders	Goal	Definition
Professional	Fiduciary duty task orientation	shareholders, investors, supervisor, customers	achieve performance goals	Leader perceives a moral obligation toward shareholders/owners, born out of fiduciary duty; displays instrumental behavior targeted at fulfilling organizational performance goals
Facilitator	Respect, care relational orientation	subordinates, colleagues	create fair working environment	Leader perceives moral obligations towards colleagues and subordinates; displays relational behavior in that she/he shows care and concern for others in the work environment
Citizen	Public welfare social welfare orientation	external stakeholders (NGOs, community, state, family, non-contractual employees, etc.)	create (long-term) value for society	Leader perceives moral obligations towards society and future generations, born out of social-welfare orientation; displays citizenship behaviors in that she/he considers social and environmental impact and emphasis long-term goals

me to directly measure responsible leadership effectiveness. I base these criteria on the multi-dimensional leadership model from Voegtlin and Schwab (2015) that was discussed in the previous part. An individual's effectiveness as a responsible leader is therefore measured by how well he or she performs in the three roles professional, facilitator, and citizen or in other words, how well he or she considers the needs of the company, the employees, and the external environment in the decision-making process. There are also two additional criteria which are not explicitly named in the multi-dimensional model, but are implicit in its fundamental assumptions. One is how well individuals fulfil all three roles simultaneously (total), the other is how well they perform as idea-provider; that is to say, how much they bring in their own ideas and creative solutions. The methods used for measuring these five criteria will be examined in Subchapter 3.2.

2.2. Characteristics

This subchapter focuses on the individual characteristics that I intend to measure within the confines of my empirical research study. I will introduce each characteristic separately, justify its inclusion into the study, and formulate a hypothesis about the way it will presumably affect responsible leadership effectiveness. Each part of this subchapter focuses on a different category of characteristics. Part 2.2. Demographic characteristics comprise all manner of socioeconomic factors in a given population, for instance income, religion, or birth rate. For the purpose of my empirical analysis, I measure three specific demographic characteristics of the study's par-

ticipants: their gender, age, and whether or not their university major is in psychology. For biological characteristics, I am interested in their right-hand second-to-fourth digit ratio and its correlation with testosterone levels. I will talk about the demographic factors as well as the sole biological factor, while Part 2.2.2 will discuss the personality factors.

2.2.1. Demographic and biological characteristics

Demographic characteristics comprise all manner of socioeconomic factors in a given population, for instance income, religion, or birth rate. For the purpose of my empirical analysis, I measure three specific demographic characteristics of the study's participants: their gender, age, and whether or not their university major is in psychology. For biological characteristics, I am interested in their right-hand second-to-fourth digit ratio and its correlation with testosterone levels.

Gender

The relationship between gender and effectiveness as a leader has been extensively studied in the existent literature, although the results are somewhat inconclusive (Eagly and Johnson (1990)). As Barbuto et al. (2007) note: "For every study that has shown differences in leadership behaviors based on gender ..., another has shown no differences at all" (p. 71). Researchers who observe gender differences in their studies often argue that these can be explained by social stereotypes (Ridgeway (2001)). Men are expected to be task-oriented; as such, their leadership style tends to be autocratic with a strong emphasis on hierarchical authority, competitiveness, and achievements (Eagly and Johnson (1990)). Meanwhile, women are expected to display a more interper-

sonal approach. Their leadership style is generally more participatory, focusing on building up morale within a team and maintaining positive relationships (Barbuto et al. (2007)). In a similar vein, some studies have found that women are more likely to score well on emotional intelligence than their male peers (Mandell and Pherwani (2003)). However, these gender differences frequently disappear when looking at a formal organisational setting (Barbuto et al. (2007)), which indicates that women in management positions tend to adopt the leadership style of men. This may be because traditional notions of leadership have been dominated by a masculine view (Carli and Eagly (2001)), pressuring women into changing their approach if they want to be seen as equals.

Since responsible leadership theories tend to emphasise the limitations of a purely task-focused approach while simultaneously stressing the importance of building relationships and cooperating with stakeholders, it seems possible that stereotypically female traits would be more valuable than they are often considered to be in a traditional understanding of leadership. Women might be more likely to build up strong relationships with employees as well as external parties, thus performing better in the roles of facilitator and citizen. However, their performance as professional is harder to assess. Shareholder obligations often require a strong focus on performing tasks as efficiently as possible, which could favour a more stereotypically masculine leadership. The literature also gives no clear indication on whether gender has an influence on the role of idea-provider. While there is a wealth of research on the relationship between gender and creativity, findings in the field have proven inconclusive (Baer and Kaufman (2008)). To summarise, while the literature offers some indication that women could potentially perform better in the roles of facilitator and citizen and men in the role of professional, there is not enough evidence to make strong assumptions about how gender differences affect the ability to perform all three roles simultaneously or to act as idea-provider.

Hypothesis 1: Women are more likely than men to perform well in the roles of facilitator and citizen, while men are more likely to perform well in the role of professional.

Age

The influence of age on leadership behaviour has not been as thoroughly explored as that of gender (Barbuto et al. (2007)). However, recent demographic changes in firms, such as the increasingly aging workforce widening the gap between the oldest and youngest employees and the greater prevalence of young managers in high positions, have spurred interest in the topic (Oshagbemi (2004)). In a study focusing on the Australian Public Service and National Health Service trust organisations in the United Kingdom, Korac-Kakabadse et al. (1998) found that time-related dimensions such as a manager's number of years in the organisation, number of years in the job, and age have a noticeable effect on his or her leadership behaviour. Young managers are more likely to be radicals who dislike control, but enjoy

challenges. They are primarily characterised by their flexibility and competitiveness as well as their independence. Meanwhile, older managers tend to be either bureaucrats who focus on rules and maintaining order or team players who adopt a more team-driven approach and encourage constructive dialogue (Korac-Kakabadse et al. (1998)). Similar findings were observed by other researchers in the field (Oshagbemi (2004)). Younger leaders generally appear to feel more comfortable in rapidly changing environments and operate with more energy, but they also tend to be more competitive and work towards promoting themselves first and foremost. Older leaders, on the other hand, try to minimise risks by relying on past experience and in-depth knowledge of their respective fields. They tend to cooperate more with their followers and primarily work towards promoting others (Oshagbemi (2004)).

Translating these findings into responsible leadership effectiveness is difficult, as both young and old leaders seem to possess beneficial qualities. The increased flexibility and drive of the younger leaders may allow them to more swiftly adapt to the increasingly complex business environment. However, their strong self-focus and competitiveness would seem to be at odds with an approach that strongly emphasises stakeholder concerns, whereas the team- and communications-oriented style of older leaders would be more suited for this task. Additionally, the increased knowledge and experience of the older leaders may potentially allow them to resolve complex and seemingly paradoxical situations better than their younger peers. As such, I would argue that age is likely to positively influence all aspects of responsible leadership effectiveness.

Hypothesis 2: Age has a positive influence on all three roles of responsible leadership (professional, facilitator, and citizen) as well as the role of idea-provider.

Psychology student

While the question of whether or not leadership can be taught is the subject of extensive debates in the field (Elmuti et al. (2005)), research that looks at the specific impact a person's education has on his or her leadership behaviour is still relatively rare (Barbuto et al. (2007)). Only a small number of empirical studies have attempted to map the causal relationship between educational level and leadership effectiveness. Similarly, there is little research into the relationship between a student's university major and his or her leadership style, despite there being evidence that students exhibit personality differences depending on their chosen field of study (Baluch et al. (1996); Vedel (2014)).

The participants of our research study were almost exclusively students from the UZH or the Swiss Federal Institute of Technology in Zurich (ETHZ) and thus all on roughly the same educational level. However, there were noticeable differences in their chosen field of study. While we had students from a wide variety of different fields, more than half of the participants stated that they were majoring in psychology. I therefore decided to create a variable for my regression

models which measures whether the responsible leadership effectiveness of psychology students differs from that of their peers.

Several empirical studies have analysed the differences in personality characteristics between students enrolled in various college or university majors. A systematic literature review by Vedel (2014) found that psychology students tend to score high on the FFM dimensions neuroticism, openness to experience, and agreeableness. Liberal arts students on average also exhibit higher creativity than those majoring in business or natural sciences (Pringle et al. (2010)). In addition, Dimitrijević et al. (2011) found that psychology majors show more empathy and a greater motivation to help others than students who have chosen non-helping professions. These results indicate that psychology students might be better at performing the roles of facilitator and citizen, as they have a greater desire to serve stakeholder needs and ensure that everyone is satisfied. Higher creativity could also mean that they perform well in the role of idea-provider.

Hypothesis 3: Students who major in psychology are more likely to perform well in the roles of facilitator, citizen, and idea-provider than students from other fields.

Originally, I had planned to include an additional variable for students who are business majors to determine whether they would differ from the other participants as well. However, only sixteen of the participants in our study actually majored in an economic field, including those who were enrolled in economics or economic chemistry. As such, I decided to discard the variable due to the low sample size.

Second-to-fourth digit ratio and testosterone

The effect of testosterone on human interactions has been extensively studied in a wide variety of medicinal and psychological research papers (Hines (2006)). Historically, the hormone has been closely linked to a variety of negative attributes, including aggressiveness, anti-social behaviour, and criminal activities. However, more recent evidence indicates that this relationship is not as clear-cut as it is commonly assumed to be (Dabbs Jr and Morris (1990)). While many studies did indeed find a correlation between aggressiveness and testosterone, others found no correlation at all (Van Bokhoven et al. (2006)). This has led some researchers to suggest that testosterone in humans is primarily linked to a desire for social dominance (Mazur and Booth (1998)). Higher levels of testosterone therefore indicate greater power motivation and increased vigilance against potential threats to one's social status (Eisenegger et al. (2011)). This inclination towards dominance can sometimes lead to aggressive or violent behaviour, but it can also manifest itself in alternate ways. For instance, Eisenegger et al. (2010) found that higher levels of testosterone increased the fairness of participants in the ultimatum bargaining game because those with high testosterone do not want their offer to be rejected by their partners. The greater degree of fairness is hereby not caused by altruism, but rather by the person's concern for his or her social status. In addition, while women on average

have far lower levels of testosterone than men, an increase in testosterone seems to have a similar effect on both genders (Archer (2006); Stanton et al. (2011)).

Based on these findings, it seems reasonable to assume that testosterone levels would have an impact on responsible leadership effectiveness as well. However, the exact nature of this relationship is difficult to discern. The study conducted by Eisenegger et al. (2010) suggests that leaders with a high testosterone level might be more willing to fulfil stakeholder needs because finding a solution that is agreeable to everyone would increase their social status and assert their dominance. At the same time, such leaders would only engage in this behaviour if it were of personal benefit to them. This could indicate that they would focus solely on powerful stakeholders who have something to offer them, such as shareholders, while neglecting stakeholders whose claims may not be any less legitimate, but who cannot reward or punish the manager. Furthermore, high-testosterone leaders might also be more likely to fall back on more aggressive and competitive tactics if the cooperative strategy does not produce the desired results. I therefore argue that higher levels of testosterone are negatively correlated with effectiveness in the roles of facilitator and citizen. However, such leaders might show a better performance in the role of professional due to their increased competitiveness and desire to improve their social standing, which in the case of a CEO is often tied to the firm's earnings.

In our research study, we measured testosterone levels directly through the use of saliva samples. However, due to financial constraints, we could only analyse samples from male participants. Additionally, providing a saliva sample was not a necessary step for partaking in the study. As a result, we only received samples from roughly half of all participants. To measure the testosterone levels of the other participants, I rely on the 2D:4D, which is the relative length between the second and fourth finger of a hand. Studies have shown that this ratio is strongly correlated with a person's prenatal testosterone levels (Neave et al. (2003)). A low 2D:4D has been linked to higher levels of prenatal testosterone and correspondingly to lower risk aversion (Coates et al. (2009)) and increased aggressiveness (Perciavalle et al. (2013)), both of which are qualities that have also been linked to current testosterone levels (Sapienza et al. (2009); Van Bokhoven et al. (2006)). In my main regression models, I include the 2D:4D to see if it influences responsible leadership effectiveness. I will also analyse the influence of the testosterone values from the saliva samples, although that will be done in a separate regression to avoid the danger of overfitting. Additionally, I will look at the correlation between the two variables to see whether prenatal testosterone values correspond to current ones. While we measured the 2D:4D for both hands in our study, I will be using the ratio from the right hand in my analysis because it tends display more robust sex differences, indicating that it is more sensitive to differences in prenatal androgens (Coates et al. (2009)).

Hypothesis 4: A person with a low 2D:4D is more

likely to perform well in the role of professional, but less likely to perform well in the roles of facilitator and citizen than one with a high 2D:4D.

Hypothesis 5: A person with high levels of testosterone is more likely to perform well in the role of professional, but less likely to perform well in the roles of facilitator and citizen than one with low levels of testosterone.

2.2.2. Personality characteristics

As has been discussed in Part 2.1.1, traditional trait theory has analysed a large variety of personality characteristics to determine whether they are correlated with effective leadership behaviour. However, researchers often worked with their own specific understanding of a given trait, making direct comparisons between studies difficult or even impossible (Judge et al. (2002)). To mitigate this issue, I based my selection of personality characteristics on validated scales which are already well-established in psychological research. My regression models include two dimensions from the HEXACO model of personality structure: honesty-humility and openness to experience. In addition, I also measure empathic concern and perspective taking from the Interpersonal Reactivity Index (IRI) as well as holism from the Analysis-Holism Scale (AHS).

Honesty-humility

One of the five dimensions of the FFM is agreeableness, which measures an individual's concern for cooperation and compliance with social norms (Jensen-Campbell and Graziano (2001)). As such, it puts a much stronger focus on interpersonal relationships than the other FFM characteristics (Graziano et al. (1996)). People who score high on agreeableness value social harmony and desire to get along well with everyone, often showing concern for the well-being of others. In social interactions, they are generally perceived to be warm, friendly, and helpful (Graziano and Eisenberg (1997)). Conversely, people who score low on agreeableness tend to be more suspicious of others and show less concern for their well-being. They are therefore also less likely to go out of their way in order to help someone and place a greater emphasis on competition rather than cooperation. Agreeableness has also been shown to correlate with empathy (Graziano et al. (2007)), indicating that those with high agreeableness display a greater degree of prosocial motivation and a higher desire to help rather than hurt.

For our research study, we used a modified version of the FFM called the HEXACO model of personality structure, which was developed by Michael Ashton and Kibeom Lee (Ashton et al. (2004)). Like the FFM, the HEXACO model incorporates the three dimensions extraversion, conscientiousness, and openness to experience, leaving them mostly unchanged. However, it replaces neuroticism and agreeableness with three new dimensions: agreeableness, emotionality, and honesty-humility. This restructuring was a reaction to a number of lexical studies which consistently found six

factors of personality rather than five, indicating that the traditional FFM approach is somewhat limited (Ashton and Lee (2007)).

In the HEXACO model, both agreeableness and honesty-humility possess elements that are associated with the traditional understanding of agreeableness and can be seen as complementary aspects (Ashton and Lee (2007)). They are both used to measure reciprocal altruism, albeit from different vantage points. Honesty-humility is an expression of fairness. Individuals who score high on this dimension tend to cooperate with others even if they have something to gain by exploiting them. Agreeableness meanwhile measures tolerance and forgiveness. A person who exhibits high agreeableness is willing to cooperate with others even if he or she has been or is currently being exploited by them (Ashton and Lee (2007)).

It seems reasonable to suggest that these two dimensions have the potential of affecting responsible leadership effectiveness. A manager who scores high on honesty-humility may be more likely to cooperate with stakeholders and consider their demands even if there is no immediate financial benefit to be gained from such behaviour. Likewise, a manager who scores high on agreeableness might be willing to persevere and keep up negotiations despite potential setbacks. The general desire for social harmony and greater willingness to help others that is inherent to these two dimensions could also be seen as positive factors when dealing with employees and external stakeholders.

Originally, I had intended to include both agreeableness and honesty-humility as separate predictors in my multiple regression models. However, I did not want to incorporate too many independent variables due to my relatively low sample size and the associated danger of overfitting. As such, I decided to discard agreeableness and instead focus solely on honesty-humility. I argue that this choice is justified because CEOs of multinational companies are frequently in a position where they can exploit their stakeholders, while stakeholders themselves often have little to no means of exploiting the company. Therefore, honesty-humility seems to be more pertinent to a manager's business activities than agreeableness and accordingly a better predictor for responsible leadership effectiveness. In conclusion, I posit that displaying a high degree of honesty-humility will positively influence a manager's ability to perform well in the roles of facilitator and citizen.

Hypothesis 6: A person who scores high on honesty-humility is more likely to perform well in the roles of facilitator and citizen than one who scores low.

Openness to experience

Like agreeableness, openness to experience is one of the personality characteristics of the FFM. It can be broadly seen as a reflection of curiosity and open-mindedness, measuring the number of interests a person has and the degree to which he or she pursues them (Howard and Howard (1995)). Those who score high on openness to experience tend to welcome and actively seek out new aspects of life. They possess a high

level of emotional sensitivity and show an active imagination as well as strong aesthetic interests (Coan (1972)). Conversely, people who score low on openness to experience are often sceptical or hostile to new practices, preferring to rely on tradition and conventional approaches instead.

Empirical research studies have found a correlation between openness to experience and general creativity (Baer and Oldham (2006); McCrae (1987)) as well as adaptability (LePine et al. (2000)). High openness to experience seems to lead to greater flexibility, making it easier to adapt to rapidly changing conditions. Likewise, people who score high on openness to experience appear to be more likely to think outside of the proverbial box and come up with new and unique solutions to existing problems. George and Zhou (2001) theorise that this is due to said people having a broader range and depth of experiences to fall back on when attempting to solve an issue as well as being more open to the idea of trying something different. Thus, they have both the means and the inclination to discover radically new approaches. In a similar vein, openness to experience has been found to have a moderate influence on general intelligence (Harris (2004)).

The HEXACO model of personality structure we use in our research study defines openness to experience in the same way as the traditional FFM, although it slightly alters the respective subcategories (Ashton and Lee (2007)). In the HEXACO model, openness to experience is comprised of four dimensions: aesthetic appreciation, inquisitiveness, creativity, and unconventionality (Lee and Ashton (n.d.)). Aesthetic appreciation assesses the degree to which an individual derives enjoyment from the beauty inherent in art and nature, inquisitiveness measures the willingness to seek out new information and experiences, creativity showcases the preference for innovation and experimentation, and unconventionality indicates how accepting an individual is of people or ideas that defy societal expectations (Lee and Ashton (n.d.)).

Leaders in a modern, globalised business world are constantly faced with rapidly changing conditions and highly complex, seemingly paradoxical situations. Hence, openness to experience might be a valuable trait for those aspiring to be responsible leaders. Managers who score high on this personality dimension may be more capable of resolving conflicting stakeholder demands by drawing upon a greater wealth of knowledge and experience as well as by attempting radically new solutions that more traditionally minded managers would shy away from. Their increased flexibility may also allow them to more swiftly resolve issues or even recognise and prevent potential threats before they can become an acute problem. I therefore argue that a high degree of openness to experience positively influences a leader's ability to perform well in the role of idea-provider.

Hypothesis 7: A person who scores high on openness to experience is more likely to perform well in the role of idea-provider than one who scores low.

Empathic concern

Empathic concern is a widely used term in psychological re-

search that refers to an emotional response of compassion and sympathy caused by perceiving another person to be in distress (Niezink et al. (2012)). It reflects an individual's intrinsic valuing of someone else's welfare (Batson (2009)) and often manifests itself in feelings of sadness, guilt, or emotional anguish. Empirical evidence suggests that this type of concern is a fundamental driver of prosocial or altruistic behaviour (Einolf (2008)), as experiencing negative emotions upon seeing another person suffer frequently causes the observer to actively try and help, if only to reduce his or her own personal level of mental distress (Schroeder et al. (1988)).

While there is some debate over the extent to which empathic concern is a fixed attribute rather than a spontaneously arising emotional reaction, some researchers do conceptualise it as a personality trait that can strongly vary between different individuals (Einolf (2008)). Empathic concern is one of the four categories measured in the IRI, where it is used to assess emotional empathy; that is to say, the degree to which a person experiences feelings of compassion and concern for an unfortunate other he or she is observing (Davis (1983)). Those who score high on empathic concern strongly feel for people in need and wish to ease their suffering, while those who score low are less affected by the well-being of others.

It seems possible that empathic concern would affect the manner in which a business leader interacts with stakeholders. Managers who display high levels of empathic concern may be more likely to feel sympathy for the plight of affected stakeholders and thus possess a greater incentive to find solutions that satisfy all parties. Conversely, managers with low empathic concern might not care as much about finding mutually beneficial solutions because the well-being of stakeholders is of lesser concern to them. Instead, they would focus their attention and efforts on satisfying their shareholders or any other stakeholders who are of direct importance to the company's survival. I therefore suggest that empathic concern is positively correlated with a leader's performance in the roles of facilitator and citizen.

Hypothesis 8: A person who scores high on empathic concern is more likely to perform well in the roles of facilitator and citizen than one who scores low.

Perspective taking

The term perspective taking refers to the cognitive ability to conceptualise situations or events from the viewpoint of another individual (Galinsky et al. (2008)). It is often considered to be an important antecedent for empathy (Oswald (2002)) and has been linked to a variety of benefits. Research on the subject indicates that possessing the capability of leaving one's own personalised frame of reference and adapting that of someone else reduces stereotyping and facilitates the exchange between in- and out-groups (Galinsky and Moskowitz (2000)) while simultaneously providing an advantage in negotiations (Galinsky et al. (2008)).

Like empathic concern, perspective taking is one of the four personality aspects measured by the IRI. While these

two dimensions are often correlated (Davis (1983)), they describe distinct concepts. Empathic concern assesses emotional empathy, which can be described as the degree to which one sympathises with another's plight. Perspective taking, on the other hand, is a purely cognitive trait, showcasing the extent to which a person knows and understands what others are thinking or feeling. It is quite possible that someone would score high on perspective taking and therefore have a pronounced understanding of another person's inner turmoil, yet at the same time not be particularly concerned with said person's well-being, thus scoring low on empathic concern.

As such, varying levels of perspective taking may not necessarily influence a manager's motivation or goals when dealing with stakeholders. However, it still seems conceivable that perspective taking could be a valuable trait in the context of responsible leadership. Being able to fully understand the thoughts and feelings of all involved parties may facilitate the negotiation process, thereby leading to improved solutions. Leaders who possess this quality could thus resolve more complex situations than their peers who lack it. Based on this line of reasoning, I hypothesise that perspective taking positively influences the ability to perform well in the role of idea-provider.

Hypothesis 9: A person who scores high on perspective taking is more likely to perform well in the role of idea-provider than one who scores low.

Holism

Holism is a concept that is used across a wide spectrum of scientific fields, although its connotations differ somewhat depending on the respective context to which it is applied (Pelletier (2012)). Generally speaking, it refers to the idea that the properties of a system can only be understood and analysed as a whole. Studying individual components in isolation will not produce meaningful insights because doing so ignores the dynamic interactions which occur among these components. Holism can thus be understood as an application of the Aristotelian principle that the whole is greater than the sum of its parts (Freeman (2005)).

The conception of holism as an inherent personality trait rather than an abstract ideal is a comparatively new phenomenon. Drawing from previous research on cultural differences, Choi et al. (2007) note that East Asians often seem to approach issues from another perspective than Westerners. While the former believe that all elements in the world are interconnected, the latter operate under the assumption that the universe is comprised of independent parts. Choi, Koo, and Choi expand on this observation by developing the AHS; a framework for measuring individual differences in analytic versus holistic thinking.

Those who score high on the AHS tend to exhibit a greater degree of holistic thinking. They focus their attention on the relationships between objects and their environment, seeing the world as a complex network of causal relationships that are dynamic and constantly changing (Choi et al. (2007)).

Subsequently, they are less likely to judge any information as irrelevant to a specific issue and report a greater connectedness to the environment (Leong et al. (2014)). When faced with contradictory demands, they are primarily concerned with preserving harmony by finding a common middle ground (Lechuga et al. (2011)). Conversely, those who score low on the AHS tend to employ analytic thinking, preferring to focus on individual objects removed from their contextual surroundings and being more willing to choose one side over the other when faced with potential disagreements.

Based on these findings, it seems possible that holistic thinking could be a beneficial attribute for responsible leaders. Managers who consider problems within their wider environmental context may be better at solving complex, interwoven issues that affect multiple stakeholder groups than those who only look at individual issues in isolation. Similarly, the fact that holistic thinkers display a greater connectedness to the environment and a strong concern for preserving harmony among different groups would presumably cause them to put more effort into analysing the needs of all stakeholders, even of those who are not vital to the company's continued operations or survival. I therefore argue that holism is positively correlated with the ability to succeed in the roles of facilitator, citizen, and idea-provider.

Hypothesis 10: A person who scores high on holism is more likely to perform well in the roles of facilitator, citizen, and idea-provider than one who scores low.

3. Methods

In Chapter 2, I derived my research question from the theoretical literature and then turned it into a set of ten quantifiable hypotheses. Chapter 3 will discuss the specific process I used for collecting the necessary data as well as the statistical methods I employ to test my hypotheses.

The main goal of this master's thesis is to investigate the causal relationship between personality characteristics and responsible leadership effectiveness by performing a quantitative empirical analysis. For this purpose, I have constructed a number of multiple linear regression models, each with one of the roles of responsible leadership as dependent variable and the various demographic, biological, and personality traits I introduced in Subchapter 2.2 as independent variables. The data for these variables was collected through surveys and a scenario-based experimental study with a sample of students from the UZH and the ETHZ as participants.

This experimental research study was a collaboration between two chairs of the UZH: the Chair of Clinical Psychology and Psychotherapy from the Department of Psychology and the Chair of Foundations of Business Administration and Theories of the Firm from the Department of Business Administration. The study was designed by Dr Christian Vögtlin and Pascale Schwab in coordination with Andreas Walther and Prof. Dr Ulrike Ehlert and supervised by Dr Voegtlin.

The following subchapters will discuss our research study in greater detail. Subchapter 3.1 concerns itself with the participants; I will discuss their selection process and analyse their composition. Subchapter 3.2 focuses on the measures used in the study, detailing all observed variables and which instruments were employed to assess them. Subchapter 3.3 is dedicated to the procedures. It shows the data collection process as well as the statistical tools I use to construct my regression models. Lastly, Subchapter 3.4 discusses the limitations of our research study.

3.1. Participants

As discussed in Subchapter 2.1, responsible leadership is a complex phenomenon with a large variety of potential influence factors. In particular, it seems likely that people's personal leadership experiences or the organisational environment under which they are operating would have a profound impact on their view of responsible leadership and thus also on their effectiveness in this regard. Since our research study is primarily concerned with assessing the relationship between personality characteristics and responsible leadership, we wanted to minimise our participants' exposure to these two variables, which is why we decided to draw a sample from a pool of university students. Many of these students are likely to have never worked in an organisation, and those who did would presumably only have worked there for a few years at most. As such, their decision-making process may not be influenced by the organisational context as much as that of someone who has worked in the same company over a long period of time. Similarly, students who worked before starting their degree or who are working while finishing their bachelor's or master's degree are less likely to have ever been put in a leadership role, which suggests that their decisions would not be strongly affected by previous leadership experiences. We thus posit that using students for measuring the relationship between personality and responsible leadership is less likely to result in the aforementioned variables distorting the results than using participants who are already in the workforce. In addition, relying on students from the UZH and the ETHZ both facilitated our recruitment process and ensured that anyone who was interested would not have to travel far to partake in the experimental part of our study.

Instead of drawing a random sample, we relied on students who were interested in the subject matter and voluntarily wished to participate. Those who completed our study had the chance of winning an iPad mini or one of two Amazon vouchers with a value of 50 CHF each, but otherwise received no financial compensation for their efforts. We primarily recruited participants through two methods: By visiting various lectures and giving a short presentation to the attending students, and by sending out information over the mailing list of the Department of Psychology. In either case, we did not reveal specific details about the nature of our research study to the respective students. Instead, we simply told them that it would centre on decision making in a business context. This was done in order to prevent the participants from approaching our study with preconceived notions

and therefore a potential bias. Students who were interested in the research study upon hearing or reading our description could give us their email address, whereupon we provided them additional information as well as a link to the online survey we used to measure personality characteristics. Afterwards, they would visit the Chair of Foundations of Business Administration and Theories of the Firm to partake in the second part of the study. These procedures will be examined in greater detail in Subchapter 3.3.

Towards the end of our data collection period, we also started employing alternative recruitment methods. On several mornings and afternoons, we directly approached random students who were sitting in the main hall of the UZH and asked them if they were interested in participating. As before, we did not reveal the exact nature of our study, instead describing it as focusing on decision making in a business context. While the recruitment process discussed in the previous paragraph almost exclusively yielded students from economics or psychology, talking to those sitting in the main hall brought in quite a few participants from other fields as well. In addition to this approach, we also recruited colleagues or friends who were not affiliated with the study. While most of these people were students, four were already in the workforce. However, none of them were employed in a leadership position, and with one exception, they all had either just started working or were only working for a few years, which is why we decided to keep them in our sample.

In total, 97 people participated in the experimental part of our study. Three of them had to be discarded from our sample because they did not fill out the corresponding personality survey, leaving us with 94 participants who completed both the survey and the experimental part. The data we obtained from these 94 men and women form the basis of my multiple regression models.

Occasionally, a participant did not receive a score for one of the personality traits we measured in the survey. This could happen because the scores are computed by averaging the values a participant gives for several questions related to the personality dimension. If he or she left one of those questions out, the total score could no longer be assessed. As a result, there are three data entries missing for holism in my model, two entries for honesty-humility, openness to experience, and empathic concern, and one entry for perspective taking. Performing Little's test of missing completely at random for all independent variables in my model produced a nonsignificant result ($\chi^2 = 23.851$ with $p = 0.736$). This indicates that the entries are missing completely at random (Little (1988)), which allows me to fill them in using the expectation-maximisation algorithm (Do and Batzoglou (2008)).

When looking at the composition of the participants, there are two notable aspects. The first is the large number of psychology students. 51 people reported that they were majoring in psychology, more than half of all participants and far more than for any other field of study. This is partly because we spent a lot of time and effort on recruiting psychology students, visiting several of their lectures and

advertising on their mailing list, which is explicitly designed to introduce research studies. However, this in itself is not sufficient to explain the large discrepancy, as we also visited quite a few economic lectures, but only had 16 participants who reported that they were majoring in that field. The disproportional amount of psychology students can largely be attributed to the fact that the Bachelor of Science in Psychology is the only major at the UZH in which the students are required to collect hourly credits as trial subjects. Every psychological study conducted by members of the university rewards its participants with a certain number of hourly credits depending on the time investment said study requires. A psychology student needs to have collected at least ten of these hourly credits before he or she is allowed to graduate. Since our research study was undertaken in collaboration with the Chair of Clinical Psychology and Psychotherapy, we were authorised to reward hourly credits as well. Thus, while students from other fields were solely motivated by interest in the subject matter or the small chance of winning a prize, psychology students also received a more tangible benefit for partaking in the study.

The second notable element is the ratio between male and female participants. Only 32 of the 94 people in our sample are women, while 62 are men. This gender discrepancy can be explained by our recruiting methods. Originally, we did not target specific individuals, instead introducing our study to large groups and then waiting for interested students to contact us. During that time period, the number of male and female participants was almost equal. However, later on, when we approached specific students sitting in the main hall, we focused almost exclusively on recruiting men. This was done because only men could deliver saliva samples, which we were then able to use for observing testosterone levels. Analysing these saliva samples is a costly process, and due to financial constraints, we knew from the start that we would not be able to do it for both genders. We therefore decided to solely analyse the samples of men because they tend to have higher testosterone levels on average (Sapienza et al. (2009)). Thus, the more men we recruited for our study, the more chances we had to obtain additional saliva samples and therefore collect more data on testosterone. We received a total of 52 samples, though I had to discard one from my analysis because we could not find the corresponding activity log. As such, I have 51 data points for the variable testosterone.

3.2. Measures

To properly assess the causal relationship between personality characteristics and responsible leadership effectiveness within the confines of a multiple linear regression model, I needed to turn my theoretical assumptions into a quantifiable set of dependent and independent variables. My independent variables are formed by the ten factors I have introduced in Subchapter 2.2: gender, age, psychology student, 2D:4D, testosterone level, honesty-humility, openness to experience, empathic concern, perspective taking, and holism.

The variables gender, age, and psychology student were measured by directly asking the participants about them. Gender and psychology student are dummy variables; for gender, 1 means male and 0 means female, while for psychology student, 1 is yes and 0 is no. Age is considered a discrete variable in my model because our survey only asked for full years. The 2D:4D was assessed when the participants personally came to the experimental part of the study. They were asked to sit down and lay their hand on their lap or a table. We then measured their second and fourth digit two times each and used the average of those values for computing the 2D:4D as a continuous variable. The testosterone level is also a continuous variable and was obtained by analysing the provided saliva samples.

The personality characteristics are all continuous variables which are computed by averaging multiple 5-point Likert questions, ranging from 1 = strongly disagree to 5 = strongly agree. Honesty-humility and openness to experience were assessed by using the 60-item version of the revised HEXACO personality inventory (Ashton and Lee (2009)), empathic concern and perspective taking by using the Saarbrücker Persönlichkeitsfragebogen, a validated German version of the IRI (Paulus (2009)), and holism through the AHS (Choi et al. (2007)). The mean and standard deviation of each variable can be found in Appendix A.

Cronbach's alpha is above 0.7 for the variables openness to experience (0.713), empathic concern (0.730), perspective taking (0.779), and holism (0.743), which is generally considered to be acceptable (Tavakol and Dennick (2011)). However, Cronbach's alpha for honesty-humility is only 0.651, indicating that this variable may be less reliable than the others. The difference is not enormous, so analysing honesty-humility may still yield useful insights, but its results should definitely be interpreted with caution.

In addition to these ten variables, our survey also assessed a variety of other personality characteristics, including the additional dimensions of the HEXACO model, internal and external loci of control, and the positive and negative affect schedule. However, I decided not to analyse those traits in this thesis partly because their relationship with responsible leadership behaviour is less apparent and partly because having too many parameters in my model would increase the possibility of overfitting. There is no clearly defined consensus in the theoretical literature as to how many predictors can be reliably measured with a given sample without the danger of overfitting (VanVoorhis and Morgan (2007)), but a commonly used rule of thumb is that for every independent variable, one should have a minimum of ten events (Harrell et al. (1996)). While I am analysing ten independent variables, I will be studying testosterone levels separately from the rest due to their low sample size and unique correlation with the 2D:4D. Therefore, my standard regression models include nine independent variables while my data set consists of 94 observations, which is in accordance with the one in ten rule.

Our selection of dependent variables was inspired by the multi-dimensional leadership model introduced in Part 2.1.3.

We decided that it would make sense to break down overall responsible leadership effectiveness into a set of continuous variables, ranging from one to five, which convey how well a participant is able to perform in each role of the leadership model. Thus, we have a dependent variable for how well someone considers the needs of the company (professional), the needs of the employees (facilitator), and the needs of the external environment (citizen). In addition, we measured how much the person in question was able to think outside the box and offer creative solutions to a problem (idea-provider) and how well he or she managed to consider all stakeholder needs simultaneously (total). Originally, we had also planned to include a sixth dependent variable which observes how well someone considers the needs of the customers, but critically evaluating this dimension revealed that it mostly measured the same elements as the variable professional because satisfying customer demands is usually seen as the best way to maximise profits. Therefore, we decided to discard it.

Observing five dependent variables does mean that the results of our study cannot be represented in a single regression model. Instead, I have created and analysed a number of distinct, yet related multiple linear regression models, each with one of the responsible leadership roles as dependent, and nine of the ten factors described in the previous paragraphs as independent variables.

The dependent variables were assessed by giving the participants a think-aloud protocol in which they had to take on the role of a CEO and decide on whether or not they would implement a new technology. We then graded their answer along the criteria outlined in the previous paragraphs. Every time they considered the needs of the company, employees, or external stakeholders, they received one point in the respective category. No negative points were given for stating that they did not care about a specific stakeholder group. At the end, we aggregated the number of points they received to form their score for each dimension, which could rank from 1 = lowest to 5 = highest. The advantage of this approach is that it allows us to analyse the participant's strengths and weaknesses separately, leading to a more nuanced portrayal than if we had simply focused on responsible leadership effectiveness as a whole. For instance, some people performed really well as professionals, but not well as citizens or vice versa. Additionally, we also gave the participants points for coming up with creative solutions that were not discussed in the protocol itself, which in the end formed their score for idea-provider. The score for how well they were able to perform all three roles simultaneously was determined by considering how they did in the three roles individually, but it was not a simple average. A participant who performed decently in all three roles would have a higher score in this category than one who performed really well in two roles, but not at all well in the remaining one.

It is fair to say that there is a certain degree of subjectivity inherent to this method, especially since participants often did not clearly structure their answer, which led to them repeating thought processes or mixing different considerations

within the same sentence, making an objective assessment more difficult. To mitigate the issue of coder bias as much as possible, the answers were rated by five people: Dr Christian Vögtlin, Pascale Schwab, and myself as well as AL and CS, two naïve coders who were otherwise not involved in the study. I will be using my own values in the multiple regression models because I am the only person who rated all 94 participants. However, I did determine the inter-rater reliability with all other coders in order to measure the extent to which we were in agreement. I relied on Cohen's kappa for this task, as it is a commonly-used instrument for assessing inter-rater reliability (McHugh (2012)). Since the categories are ordered and the degree of disagreement plays a role, I focused on measuring the weighted kappa, using both linear and quadratic weighting. In both cases, I measured the inter-rater reliability with and without the dimension total. All calculations were performed by using the statistical tools available on the website VassarStats (Lowry (n.d.)). The full results of my analysis can be seen in Table 2 below.

While there is no universal consensus on what kappa values are considered good, a common interpretation is that values between 0.61 and 0.80 constitute a substantial agreement and values between 0.81 and 0.99 an almost perfect agreement (Viera et al. (2005)). Under this definition, my results are reassuring. The lowest observed value is 0.688, which still falls under the notion of substantial agreement. Several of the observed kappa values are above 0.8 or even 0.9, indicating that inter-rater reliability is very high when measuring the responsible leadership roles and that individual biases therefore did not distort the findings to a large extent. However, it should be pointed out that Pascale Schwab only rated the first 38 participants. As such, the reliability between her values and my own has to be analysed with caution.

There are two additional factors worth mentioning. Firstly, the kappa is always lower when the variable total is included, indicating that the inter-rater reliability for this variable is lower than it is for the others on average. This can be explained by the fact that total has less clear-cut criteria by which it could be measured, making it harder to objectively assess. Coders had to rate participants' performance over all three roles, which is more challenging than rating their performance for one individual role. Secondly, the inter-rater reliability between my values and those of the other researchers is higher than the one between my values and those of the two naïve coders. This is presumably because Christian Voegtlin, Pascale Schwab, and I were closely working together on the study and had read the same theoretical literature on responsible leadership, so we had a similar understanding of the concept and how to measure it. By contrast, the naïve coders were not familiar with the literature on responsible leadership and graded solely based on our instructions, so their conception of various terms may have differed from ours in certain respects.

Table 2: Inter-rater reliability between the values of other coders and my own

Coders	Observed Cohen's kappa			
	Linear without total	Linear with total	Quadratic without total	Quadratic with total
AL	0.715	0.688	0.831	0.816
CS	0.725	0.718	0.845	0.838
Voegtlin	0.825	0.806	0.903	0.892
Schwab	0.840	0.831	0.913	0.908

3.3. Procedures

The data gathering process occurred over a time period of several months, but with multiple breaks in-between. The first 43 participants were recruited and observed from the 15th of April 2015 to the 24th of June 2015, while the second set, consisting of 42 people, participated between the 22nd of October 2015 and the 18th of December 2015. Finally, the last nine participants joined our study after the New Year, in the time period between the 13th and 22nd of January 2016.

Everyone who wished to participate in our research study had to start by filling out the corresponding personality survey, which was hosted on the online platform Unipark (www.unipark.com). The full storybook, which includes all survey questions, can be found in Appendix B. In order to protect their confidentiality, the participants were asked to choose an alias at the start of the survey, which they would then keep for the remainder of the study. Once the survey was completed, they would make an appointment with us and come to the Chair of Foundations of Business Administration and Theories of the Firm for the experimental part of our study. While we tried to plan the appointments in such a way that there would always be two to four people at any given date, short-term cancellations meant that we occasionally only had one person show up on a particular day.

Regardless of the number of participants on a given day, the study always proceeded in the same manner. Once the people arrived, they were asked to sign a consent form, which can be found in Appendix C. Those who had volunteered to provide a saliva sample were then given the opportunity to do so. Afterwards, all participants filled out a mood scale assessing their current frame of mind as well as a short questionnaire focusing on the well-known trolley problem (Kamm (1989)). Once they were finished with these two items, we led them into a different room where we measured a number of biometrical factors: their height, weight, blood pressure, and the length of their second and fourth digit of both the left and right hand, which we later used to compute the 2D:4D. We also took a picture of each participant.

In the second stage of the experimental part, the participants were isolated from each other and given a think-aloud protocol in which they were asked to assume the role of a CEO of a multinational company. The full protocol is replicated below:

Sie sind der CEO eines multinationalen Unternehmens in der Energiebranche mit Standort in der Schweiz. Sie müssen darüber entscheiden, ob Sie eine neue Technologie zur Förderung von Erdöl in ihrem Unternehmen einführen wollen oder nicht. Die neue Technologie ermöglicht eine höhere Fördermenge und eine Kostensenkung, da durch die Technologie weniger Personal benötigt wird. Doch die neue Technologie ist umstritten, da sie schädlicher für die Umwelt ist als die bisherige Technologie. Wie möchten sie, unter Berücksichtigung der verschiedenen Stakeholder (z.B. die Mitarbeiter, Kunden, Shareholder und die Gesellschaft), weiter vorgehen?

In English:

You are the CEO of a multinational company in the energy sector headquartered in Switzerland. You have to decide whether you wish to implement a new technology for the extraction of crude oil in your company or not. The new technology would increase the extraction capacity and lead to a reduction in costs because fewer employees are needed for it. However, the new technology is controversial because it is worse for the environment than conventional technologies. How would you proceed when taking into account the various stakeholders (for instance, employees, customers, shareholders, and society)?

Once they had finished reading the protocol, the participants were given the opportunity to formulate and justify their own preferred strategy. We did not interfere with their thought process, nor did we set a time limit or specific formal requirements. We simply recorded their answers on tape so that we could later analyse how strongly they considered each stakeholder group and how much they brought in their own unique solutions to the problem. Both their final decision and the arguments they brought forward while reaching it were coded in accordance with the criteria outlined in the previous subchapter.

If more than one participant was present on that particular day, we also conducted a group discussion. The par-

ticipants were once again asked to assume the role of leaders in a company who have to make an important decision that affects numerous stakeholders. As with the individual think-aloud protocol, their discussion was graded along the dimensions of professional, facilitator, citizen, idea-provider, and total. In addition, we also assessed how strongly a participant dominated the discourse. While this group exercise yielded interesting insights in its own right, I decided not to include it in my thesis for two reasons. Firstly, not all participants were able to partake in one of these discussions, leaving us with fewer data points to analyse. Secondly, there are many factors which are not present in an individual assignment, but play an important role in group interactions. Properly analysing them would require setting up new hypotheses and variables, which would have been outside the scope of this master's thesis.

Originally, we had intended that the participants would grade each other's performance in the group discussion as well. However, we soon realised that they were often hesitant to give their partners a bad evaluation, leading to artificially inflated scores. As such, we decided to discard this variable halfway through the study. From that point onwards, the only thing participants had left to do after the group discussion was answering another short personality survey. This one wasn't directly related to responsible leadership, rather it was requested by a student of the Chair of Clinical Psychology and Psychotherapy for her own master's thesis. Finally, participants had the option of asking us to send them their results of the online survey. The feedback form they received is replicated in Appendix D.

Once all the data from both the initial personality survey and the experimental part had been collected, we combined it into a single file using the statistical software SPSS Statistics. I have also relied on SPSS for constructing and evaluating my multiple regression models, which will be discussed in-depth in Chapter 4.

3.4. Limitations

Our research study is one of the first attempts to empirically assess responsible leadership effectiveness. While we believe that it generally succeeds at this task, there are a number of limitations that need to be addressed. Some of them are deliberately imposed delimitations, while others are unintended side effects inherent to the study's design. In the following paragraphs, I will first discuss our specific delimitations. Afterwards, I will analyse the study's limitations and their potential influence on the observed variables while simultaneously offering suggestions as to how they can be avoided in future research.

In our study, we were solely interested in the relationship between responsible leadership and personality characteristics. As such, relying almost exclusively on students from the UZH and the ETHZ should not meaningfully distort our results. The fact that all participants joined us voluntarily, rather than being drawn from a random sample, should not have any adverse effects either because we were not trying to create generalised observations about the population of

university students. However, this delimitation does mean that we cannot assess other factors that are generally considered to have an influence on leadership effectiveness, such as the organisational culture or a person's experience in the workforce. Additionally, focusing solely on traits neglects the important role different behavioural inputs and situational contingencies play in the leadership process. The results of our study can therefore never fully explain responsible leadership effectiveness. Instead, our intention was to provide valuable insights by analysing one single dimension of a complex, multi-faceted phenomenon.

Another delimitation lies in our method for assessing responsible leadership effectiveness. In Part 2.1.2, I presented two somewhat differing definitions of responsible leadership. The one from Waldman (2011) argued that being responsible simply requires a leader to carefully consider the needs of all stakeholders, while that of Maak and Pless (2006) posited that this in itself is not enough and that responsible leadership requires actively involving the affected stakeholders in the decision-making process. Our study is founded upon Waldman's definition, as we simply analysed the extent to which participants considered the various stakeholder groups in their decision. We chose to go with this understanding partly because it makes it easier to analyse the relationship between personality and leadership effectiveness in isolation from other factors and partly because it seems to be a more accurate reflection of the leadership process as it is currently being practiced in most multinational companies. That being said, it would certainly be interesting to see a study which focuses on the direct interactions between CEOs and their stakeholders and whether a close interaction between the two could facilitate responsible leadership effectiveness, and I hope to see future studies with a focus on this aspect.

In addition to these delimitations, there are a number of unintended factors which may inhibit the validity of our results. Our personality characteristics were all measured through self-assessments. The participants were asked a number of questions, and their answers in turn determined their score for a given personality dimension. Such self-assessments can lead to distorted results if the participants do not answer all questions truthfully, which can happen either intentionally if they want to present themselves in a manner that they deem to be socially desirable or unintentionally if the person in question did not read the survey carefully or has an unrealistic and inflated view of him- or herself (McDonald (2008)). In both cases, the participant's score on a personality scale would not match his or her actual personality. This issue is further exacerbated by the fact that we used two differing instruments to measure our independent and dependent variables and also assessed them on different days, making it more difficult to directly observe the relationship between personality traits and responsible leadership. While relying on validated scales should help mitigate the problem to an extent (McDonald (2008)), it is still something that has to be kept in mind when analysing the final results.

A similar problem arises when looking at the answers

given in the think-aloud protocol. While these answers were graded by us and therefore are not subject to the exact same biases as they would have been in a self-assessment, there is still some degree of uncertainty over how well they reflect a participant's actual views. Some participants may have deliberately chosen a solution to the think-aloud protocol that they deem to be socially acceptable, rather than one they would have selected if they had not been directly observed by us. In other words, there exists a potential risk that instead of giving their honest opinion on the matter, certain participants came up with answers which they thought we wanted to hear. Allowing a participant to answer the protocol isolated from any observers on a computer may be a possible remedy for this problem, although it could also be argued that this behaviour is not actually in itself a distortion of the leadership process. After all, CEOs of multinational companies are constantly facing social pressure. The willingness to change one's approach depending on what one considers socially desirable may be a useful trait to have in terms of attaining responsible leadership effectiveness.

There is also the question of how accurately our think-aloud protocol reflects the actual leadership process. Financial and temporal constraints restricted us from constructing an incredibly elaborate role-playing scenario, making it more difficult to realistically model decision making in a business context. Our participants had to choose a suitable strategy all by themselves in a very short period of time. They had no means of collecting additional information or consulting with employees or external advisors. By contrast, the implementation of a ground-breaking new technology in an actual company is frequently a long process during which various opinions and facts are collected and processed. Furthermore, our participants received no material compensation regardless of their final decision. In a real-life scenario, managers would presumably be faced with the prospect of different rewards depending on what strategy they choose to pursue, potentially skewing their priorities. A future study which responds to these limitations by expanding the scope of leadership activities participants can undertake and incorporating a reward structure that incentivises different behaviours would be a valuable contribution to the field.

Even if we operate under the assumption that the think-aloud protocol can be seen as a useful abstraction of the leadership process, there is still the risk that its specific wording influences the outcome. While we attempted to keep the protocol's description as neutral as possible so as not to prematurely influence the reader's opinion, I cannot entirely rule out the possibility that its formulations may have affected the participants in some form. Even seemingly minor and subtle differences in the phrasing could have a subconscious impact on them. For instance, the text of our protocol could be interpreted as suggesting that the central conflict lies in the additional profits generated from implementing the new technology versus the harm done to the environment. This could potentially lead to participants focusing strongly on these two dimensions while neglecting the employees. While this in itself is not necessarily a huge issue, it

does raise concerns about the replicability of the study's results. Another study using its own decision-making scenario may reach different conclusions by unintentionally steering the participants' thoughts into a certain direction through the way its central problem is framed. I thus have to be careful in my analysis. Not all observable results may have been caused by the causal relationship between personality and responsible leadership. Some could be predicated on the wording of the presented leadership problem.

One final aspect that has to be mentioned is our assessment of responsible leadership effectiveness. This is the first research study using the five dimensions professional, facilitator, citizen, idea-provider, and total to model responsible leadership. While we believe that our justification for selecting and grading these five dimensions is sound, they have not yet been validated in previous research. As such, I cannot entirely exclude the possibility that they are an inadequate instrument for accurately measuring responsible leadership effectiveness. The high inter-rater reliability between my own scores and those of the other coders is somewhat reassuring, as it indicates that the results are not strongly influenced by subjective biases. However, it is still possible to imagine that there exist conceptual issues with the scales. Responsible leadership is a complex phenomenon and trying to dismantle it into a set of distinguishable elements a challenging task, which is why I would welcome it if additional studies used our scales to aid in further calibrating them as well as increasing their accuracy and precision.

4. Results

In Chapter 3, I explained the methodology I used to collect my data. Chapter 4 will now present the actual results of my research. I will introduce each multiple linear regression model I have constructed separately and showcase its outcomes. The in-depth interpretation and analysis of my results will be the focus of Chapter 5.

Each subchapter of this chapter focuses on one of the five dependent variables I have previously introduced. Subchapter 4.1 will discuss the variable for professional, 4.2 the one for facilitator, 4.3 the one for citizen, 4.4 the one for idea-provider, and 4.5 the one for total. For each variable, I have constructed two regression models; one in which I include all nine independent variables, and one in which I only include the independent variables for which I hypothesised that they would have an influence on the respective dependent variable. I will then compare these two models to see which one produces more significant results. Whenever applicable, I have also constructed a third model which only includes variables from the first two models that were statistically significant.

For the reasons outlined in Subchapter 3.2, I did not want to include testosterone as an independent variable in my main regression models. However, I was still curious to see whether the hormone would have a noticeable effect on responsible leadership effectiveness and how it relates to the

2D:4D. Hence, I will be analysing testosterone separately in Subchapter 4.6.

4.1. Professional

This subchapter focuses on the dependent variable professional, which measures how strongly an individual considered the needs of the company in his or her decision-making process. I will first show the model in which all nine independent variables are included, henceforth referred to as model P1. It has a Durbin-Watson statistic of 2.068, which is very close to 2. Thus, we can reasonably assume that it suffers from neither positive nor negative autocorrelation and that the error terms are independent of each other (Durbin and Watson (1951)). Furthermore, looking at the skewness and kurtosis of the of the error terms reveals that their absolute value is less than twice the standard error, indicating that the error terms are approximately normally distributed (Adams and Lawrence (2014)). Performing the Breusch-Pagan test returns a p value of 0.582, which means I cannot reject the null hypothesis that the independent variables are homoscedastic for a significance level of $\alpha = 0.05$ (Breusch and Pagan (1979)). In other words, I can operate under the assumption that no heteroscedasticity is present in my model. Table 3 below shows the findings for P1, including R , R^2 , the adjusted (adj.) R^2 , the standard error of the estimate (SEE), the F value, and the p value for the F -test.

P1 has an extremely low R^2 of 0.019, which suggests that only 1.9% of the variation in the dependent variable can be explained by the independent variables. The adjusted R^2 is even negative, which for the purpose of interpretation can be seen as equivalent to an adjusted R^2 of 0. Furthermore, the p value of the F -test is 0.996, meaning I cannot reject the null hypothesis that the intercept-only model is equal to my model for any reasonable significance level. Table 4 below shows the results for the individual variables, including the unstandardised regression coefficient b , the standard (std.) error, the standardised coefficient beta, the p value, and the variance inflation factor (VIF).

The VIF for all coefficients lies between 1.0 and 1.4, indicating that the model has no issues with multicollinearity (Alin (2010)). However, none of the included variables are statistically significant for an α of 0.05, as even the lowest observed p value is still above 0.6. As such, I cannot infer that any of the tested coefficients meaningfully influence the variable professional. These findings combined with the results from Table 3 indicate that the model P1 is a poor fit for the observed data.

In my second model, henceforth referred to as P2, I only included the independent variables for which I hypothesised that they would affect a leader's ability to perform well in the role of professional: gender, age, and the 2D:4D. The Durbin-Watson statistic for P2 is 2.034, indicating that there is no autocorrelation. The error terms are normally distributed, and the Breusch-Pagan test shows no signs of heteroscedasticity ($p = 0.486$). The results for the model as a whole and the individual coefficients are shown in Tables 5 and 6 below.

R^2 has decreased slightly, which is to be expected given that I have reduced the number of independent variables in the model. The adjusted R^2 and the p value of the F -test are slightly better than they were in P1, but still not satisfactory by any means. The same can be said about the individual coefficients. While the VIFs are all very low, and multicollinearity is thus presumably not a problem, neither of the three measured variables is significant for an α of 0.05. In conclusion, while the model P2 seems to be a small improvement over P1, it still is in no way a good fit for the data and lacks explanatory power. I thus find no support for my hypotheses that a leader's gender, age, and 2D:4D have any influence on his or her ability to perform well in the role of professional.

4.2. Facilitator

This subchapter discusses the models with the dependent variable facilitator, which assesses how strongly an individual considered the needs of the employees in his or her decision-making process. I will first introduce the model F1, which includes all nine independent variables. The Durbin-Watson statistic is 1.732, which is not as good as it was for the models P1 and P2, but still relatively close to 2. At most, F1 might display some very slight positive autocorrelation. Looking at the skewness and kurtosis indicates that the error terms are approximately normally distributed. The Breusch-Pagan test gives a p value of 0.222, which suggests that no heteroscedasticity is present. The findings of F1 are shown in Table 7 below.

R^2 is 0.213, which indicates that approximately 21% of the variation in the variable facilitator can be explained by the independent variables. Furthermore, the p value of the F -test is significant for an α of 0.05, which means that F1 is a better fit for the data than an intercept-only model would be. Table 8 below shows the results for the individual coefficients.

The VIFs are all relatively low, indicating that there is no multicollinearity in the model. From the p values, we can see that empathic concern and openness to experience are statistically significant for an α of 0.05. Empathic concern has a positive effect on facilitator; if the former increases by one standard deviation, the latter increases by 0.407 standard deviations. Conversely, facilitator's relationship with openness to experience is negative; if openness to experience increases by one standard deviation, facilitator decreases by 0.246 standard deviations.

The second model, F2, includes only the variables for which I hypothesised they would influence the role of facilitator. As such, I took out openness to experience and perspective taking. The Durbin-Watson statistic is 1.799. The error terms display a very slight positive skewness, indicating they might be asymmetrically distributed. Therefore, I used the White test for determining heteroscedasticity (White (1980)), rather than the Breusch-Pagan. Performing the White test returned a p value of 0.197, so the assumption of homoscedasticity seems to hold. The results for F2 are summarised in Tables 9 and 10 below.

Table 3: Summarised results for P1

R	R ²	Adj. R ²	SEE	F value	p value
0.136	0.019	-0.87	1.189	0.177	0.996

Table 4: Individual coefficients of P1

	Coefficient b	Std. Error	Beta	p value	VIF
(Constant)	4.658	4.350		0.287	
Gender	0.042	0.298	0.018	0.888	1.331
Age	-0.012	0.031	-0.044	0.700	1.090
Psychology	-0.149	0.285	-0.065	0.603	1.344
Honesty	0.110	0.229	0.055	0.631	1.115
Openness	-0.093	0.204	-0.050	0.652	1.055
Empathic	0.057	0.193	0.035	0.768	1.175
Perspective	0.082	0.218	0.046	0.707	1.246
Holism	-0.024	0.365	-0.008	0.949	1.186
2D:4D	-1.923	4.005	-0.054	0.632	1.098

Table 5: Summarised results for P2

R	R ²	Adj. R ²	SEE	F value	p value
0.079	0.006	-0.027	1.156	0.188	0.905

Table 6: Individual coefficients of P2

	Coefficient b	Std. Error	Beta	p value	VIF
(Constant)	5.421	3.817		0.159	
Gender	0.056	0.255	0.023	0.828	1.025
Age	-0.006	0.029	-0.023	0.827	1.005
2D:4D	-2.419	3.756	-0.068	0.521	1.022

Table 7: Summarised results for F1

R	R ²	Adj. R ²	SEE	F value	p value
0.462	0.213	0.129	0.849	2.530	0.013

Both the R^2 and the adjusted R^2 are smaller than in F1, which makes sense since F2 does not include the statistically significant variable openness to experience. The p value for the F-test is also slightly higher, though still significant for an alpha of 0.05. No multicollinearity seems to be present. The variable empathic concern is still highly significant, though its effect on facilitator has decreased slightly from 0.407 to 0.365. In general, it seems like F1 is a better fit for the data than F2. For the sake of completeness, I also constructed a model F3, which only includes the variables from F1 that were shown to be significant: empathic concern and openness to experience. The Durbin-Watson statistic for F3 is

1.788, and the error terms are normally distributed. However, the Breusch-Pagan test now returns a p value of 0.052, which is just barely above the threshold of 0.05. This suggests that F3 should be interpreted with caution, as slight heteroscedasticity may be present in the model. The results can be found in Tables 11 and 12 below.

F3 has the highest adjusted R^2 of all three facilitator models because it only includes significant variables. Similarly, it has the lowest p value for the F-test and the highest F value. When looking at the individual coefficients, the results are similar to F1. Both empathic concern and openness to experience are significant for an alpha of 0.05. The former has a

Table 8: Individual coefficients of F1

	Coefficient b	Std. Error	Beta	p value	VIF
(Constant)	0.731	3.107		0.814	
Gender	0.310	0.213	0.162	0.150	1.331
Age	-0.002	0.022	-0.009	0.930	1.090
Psychology	0.257	0.204	0.142	0.210	1.344
Honesty	0.084	0.163	0.052	0.609	1.115
Openness	-0.362	0.146	-0.246	0.015	1.055
Empathic	0.534	0.138	0.407	0.000	1.175
Perspective	-0.090	0.155	-0.062	0.565	1.246
Holism	-0.171	0.261	-0.069	0.513	1.186
2D:4D	1.032	2.861	0.037	0.719	1.098

Table 9: Summarised results for F2

R	R ²	Adj. R ²	SEE	F value	p value
0.387	0.150	0.080	0.872	2.163	0.045

Table 10: Individual coefficients of F2

	Coefficient b	Std. Error	Beta	p value	VIF
(Constant)	-0.760	3.131		0.809	
Gender	0.221	0.216	0.116	0.309	1.297
Age	-0.008	0.023	-0.034	0.739	1.067
Psychology	0.186	0.207	0.102	0.371	1.313
Honesty	0.069	0.167	0.043	0.681	1.105
Empathic	0.480	0.139	0.365	0.001	1.130
Holism	-0.212	0.253	-0.085	0.405	1.054
2D:4D	1.478	2.922	0.052	0.614	1.085

Table 11: Summarised results for F3

R	R ²	Adj. R ²	SEE	F value	p value
0.421	0.178	0.159	0.834	9.822	0.000

Table 12: Individual coefficients of F3

	Coefficient b	Std. Error	Beta	p value	VIF
(Constant)	1.395	0.658		0.037	
Openness	-0.333	0.140	-0.227	0.020	1.011
Empathic	0.498	0.126	0.379	0.000	1.011

positive and the latter a negative influence on the dependent variable facilitator. I thus find support for the hypothesis that empathic concern positively affects a leader's ability to perform well in the role of facilitator, though I cannot say the same for gender, age, psychology student, honesty-humility,

holism, or the 2D:4D.

4.3. Citizen

This subchapter is centred on the dependent variable citizen, which measures how strongly an individual considered

the needs of the external stakeholders and the environment in his or her decision-making process. The first model, C1, includes all nine independent variables. Its Durbin-Watson statistic is 2.027, which suggests that there is no autocorrelation. The error terms are normally distributed, and the Breusch-Pagan test gives a p value of 0.529, indicating that no heteroscedasticity is present. The general findings of the model are summarised in Table 13 below.

R^2 is 0.111 which indicates that 11.1% of the variation in the dependent variable citizen can be explained by the nine independent variables. The p value for the F-test is higher than the significance level of 0.05. Therefore, I once again cannot reject the null hypothesis that C1 is equal in explanatory power to an intercept-only model. These values suggest that C1 is not a good fit for the observed data. The results for the individual coefficients are shown in Table 14 below.

The low VIFs indicate that no multicollinearity is present. Of the nine observed variables, holism is significant with a p value of 0.007. The relationship between it and citizen appears to be positive; if holism increases by one standard deviation, citizen increases by 0.309 standard deviations. Thus, C1 hints at the possibility that holism might have a positive influence on the role of citizen.

The model C2 includes only the variables for which I hypothesised they would directly influence the role of citizen. As with the role of facilitator, this includes every variable except openness to experience and perspective taking. The Durbin-Watson statistic is 2.026, and the error terms are approximately normally distributed. The p value for the Breusch-Pagan test is 0.336, indicating that no heteroscedasticity is present. The summarised results for C2 can be seen in Tables 15 and 16 below.

The adjusted R^2 is slightly higher than it was for C1, although it is still relatively low. In the same vein, while the p value of the F-test is lower than in C1, it still does not meet the threshold of 0.05. Thus, I cannot confidently say that C2 is better than an intercept-only model. The low VIF of all coefficients suggest no problems with multicollinearity, but none of previously nonsignificant variables have become significant. Once again, the p value for holism is considerably lower than for any other predictor and statistically significant for an alpha of 0.05, although the p value is higher in this model than it was in C1. Looking at the bivariate correlations shows that perspective taking and holism are somewhat correlated ($r = 0.351$ with $p = 0.001$), which could explain why dropping the variable perspective taking from the model would increase the p value of holism (the full correlation table can be found in Appendix A). Since there was only a single significant variable, I decided not to construct an additional model C3. In conclusion, both C1 and C2 seem to support the hypothesis that holism has a positive influence on the role of citizen.

4.4. Idea-provider

In this subchapter, I will present the results for the dependent variable idea-provider, which shows how strongly

an individual brought in his or her own unique ideas and creative solutions to the presented problem. Model I1 includes all nine independent variables. Its Durbin-Watson statistic is 1.931. The error terms display a slight positive skewness, which once again means that I cannot measure heteroscedasticity with the Breusch-Pagan test. Performing the White test instead returned a p value of 0.090, which means I cannot reject the null hypothesis that the model is homoscedastic for an alpha of 0.05. Table 17 below shows the results for I1.

R^2 is 0.185, indicating that 18.5% of the variation in the variable idea-provider can be explained through the independent variables. Furthermore, the F-test is significant for an alpha of 0.05, indicating that this model is not a bad fit for the data. The results for the individual coefficients are summarised in Table 18 below.

The VIFs show no sign of multicollinearity. Of the nine observed independent variables, three appear to be statistically significant for an alpha of 0.05: gender, psychology student, and perspective taking. Gender and psychology student are dummy variables and have a negative effect on the dependent variable. The score for men in the role of idea-provider was on average 0.519 points lower than for women. Similarly, the score for psychology students was 0.556 points lower on average than that of the other participants. Conversely, perspective taking has a positive influence on idea-provider; if perspective taking increases by one standard deviation, idea-provider increases by 0.235 standard deviations.

The second model, I2, only includes the independent variables for which I argued that they would have an impact on a leader's ability to perform well in the role of idea-provider: age, psychology student, openness to experience, perspective taking, and holism. The Durbin-Watson statistic is 1.827, and the error terms show a noticeable positive skewness. The p value of the White test is 0.688, indicating that no heteroscedasticity is present. The full results for I2 can be seen in the Tables 19 and 20 below.

The R^2 and adjusted R^2 are considerably lower than in I2, and the F-test no longer produces a significant outcome. This can be explained by the fact that the statistically significant variable gender is no longer included in the model. Looking at the individual coefficients shows that psychology student and perspective taking are still significant, while all other variables remain nonsignificant. The beta for perspective taking has barely changed from I1, but the influence of psychology student on idea-provider is now noticeably smaller. This may be because gender and psychology student have a relatively high negative correlation ($r = -0.434$ with $p = 0.000$), so the variable psychology student ends up partly compensating for the influence of gender. Given all these factors, I1 is definitely the more fitting model for this data set with a higher explanatory power.

Lastly, I once again constructed a model, I3, incorporating only the significant independent variables: gender, psychology student, and perspective taking. The Durbin-Watson statistic is 1.959, and the error terms display a positive skewness. The White test shows no signs of heteroscedasticity (p

Table 13: Summarised results for C1

R	R ²	Adj. R ²	SEE	F value	p value
0.333	0.111	0.016	0.934	1.165	0.328

Table 14: Individual coefficients of C1

	Coefficient b	Std. Error	Beta	p value	VIF
(Constant)	-1.162	3.419		0.735	
Gender	0.177	0.235	0.089	0.454	1.331
Age	0.028	0.025	0.120	0.268	1.090
Psychology	0.117	0.224	0.062	0.604	1.344
Honesty	-0.010	0.180	-0.006	0.957	1.115
Openness	-0.044	0.161	-0.029	0.783	1.055
Empathic	0.118	0.152	0.087	0.439	1.175
Perspective	-0.188	0.171	-0.127	0.274	1.246
Holism	0.793	0.287	0.309	0.007	1.186
2D:4D	0.824	3.148	0.028	0.794	1.098

Table 15: Summarised results for C2

R	R ²	Adj. R ²	SEE	F value	p value
0.311	0.097	0.023	0.931	1.313	0.254

Table 16: Individual coefficients of C2

	Coefficient b	Std. Error	Beta	p value	VIF
(Constant)	-1.775	3.340		0.596	
Gender	0.156	0.231	0.079	0.500	1.297
Age	0.023	0.024	0.102	0.339	1.067
Psychology	0.083	0.221	0.044	0.709	1.313
Honesty	-0.029	0.178	-0.018	0.869	1.105
Empathic	0.085	0.148	0.063	0.566	1.130
Holism	0.689	0.270	0.269	0.012	1.054
2D:4D	1.228	3.117	0.042	0.695	1.085

Table 17: Summarised results for I1

R	R ²	Adj. R ²	SEE	F value	p value
0.430	0.185	0.097	0.853	2.113	0.037

= 0.158). Tables 21 and 22 below show the full results for model I3.

The adjusted R^2 is slightly higher than it was in I1, as I have eliminated the nonsignificant variables. The F value is also higher, while the p value for the F-test is lower. The individual coefficients do not reveal any new insights though. All three independent variables are still statistically significant,

and while their exact value for beta has changed slightly, their general impact on idea-provider is still the same. The models I1 to I3 thus all seem to indicate that women, non-psychology students, and those who score high on the personality dimension perspective taking perform better in the role of idea-provider than their respective counterparts.

Table 18: Individual coefficients of I1

	Coefficient b	Std. Error	Beta	p value	VIF
(Constant)	1.621	3.120		0.605	
Gender	-0.519	0.214	-0.275	0.018	1.331
Age	-0.010	0.023	-0.047	0.648	1.090
Psychology	-0.556	0.205	-0.311	0.008	1.344
Honesty	0.249	0.164	0.158	0.132	1.115
Openness	0.091	0.147	0.063	0.534	1.055
Empathic	-0.171	0.138	-0.132	0.220	1.175
Perspective	0.333	0.156	0.235	0.036	1.246
Holism	-0.241	0.262	-0.099	0.361	1.186
2D:4D	0.234	2.873	0.008	0.935	1.098

Table 19: Summarised results for I2

R	R ²	Adj. R ²	SEE	F value	p value
0.299	0.090	0.038	0.880	1.730	0.136

Table 20: Individual coefficients of I2

	Coefficient b	Std. Error	Beta	p value	VIF
(Constant)	1.720	1.190		0.152	
Age	-0.002	0.023	-0.011	0.916	1.058
Psychology	-0.397	0.188	-0.221	0.037	1.061
Openness	0.025	0.149	0.017	0.869	1.021
Perspective	0.335	0.157	0.236	0.036	1.186
Holism	-0.243	0.266	-0.099	0.364	1.150

Table 21: Summarised results for I3

R	R ²	Adj. R ²	SEE	F value	p value
0.368	0.135	0.106	0.848	4.693	0.004

Table 22: Individual coefficients of I3

	Coefficient b	Std. Error	Beta	p value	VIF
(Constant)	1.505	0.587		0.012	
Gender	-0.490	0.205	-0.260	0.019	1.233
Psychology	-0.602	0.196	-0.336	0.003	1.244
Perspective	0.282	0.140	0.199	0.047	1.014

4.5. Total

The last dependent variable I analyse within the confines of this thesis is total, which assesses how well an individual performed all three roles of responsible leadership simultaneously. The first model, T1, includes all nine independent variables. Its Durbin-Watson statistic shows no signs of au-

tocorrelation (2.078), and its error terms are approximately normally distributed. The Breusch-Pagan test has a *p* value of 0.150, which is not enough to reject the null hypothesis for an alpha of 0.05. The results for T1 are summarised in Table 23 below.

R^2 is 0.095, which means that roughly 9.5% of the vari-

Table 23: Summarised results for T1

R	R ²	Adj. R ²	SEE	F value	p value
0.308	0.095	-0.002	1.024	0.977	0.465

ation in the dependent variable total can be explained by the independent variables. However, the adjusted R^2 is less than zero, indicating that this result comes primarily from the large number of variables in the model, rather than said variables possessing any kind of explanatory power. Similarly, the p value of the F-test is larger than 0.05; therefore, I cannot reject the null hypothesis that the intercept-only model would be equal to T1. As for the individual coefficients, their results are shown in Table 24 below.

Empathic concern is the only significant variable for an alpha of 0.05. The relationship is positive; if empathic concern increases by one standard deviation, total increases by 0.247 standard deviations. This appears reasonable given that empathic concern had a noticeable effect on a leader's ability to perform well in the role of facilitator. This would also explain why openness to experience is close to being significant, though oddly enough, the p value for holism is very high despite it being statistically significant in the models C1 and C2.

In contrast to the other four dimensions of responsible leadership, I did not create any specific hypotheses as to which traits would affect a leader's ability to perform all three roles simultaneously. However, I was curious to see whether openness to experience would become significant if I were to remove some of the nonsignificant variables from T1. Hence, I created a second model, T2, which only includes the two independent variables empathic concern and openness to experience. T2's Durbin-Watson statistic is 2.153, showing no signs of autocorrelation. The error terms are approximately normally distributed, and the Breusch-Pagan test gives a p value of 0.207. Tables 25 and 26 below show the full results.

While the R^2 is slightly lower than it was in T1, the adjusted R^2 is higher due to the removal of several nonsignificant variables. Additionally, the F-test now becomes significant for an alpha of 0.05. The low VIFs indicate that no multicollinearity is present. However, while the p value of openness to experience is slightly lower than it was in T1, it still does not pass the threshold. Empathic concern shows no strong differences compared to the previous model; its p value, standard error, and beta are almost the same as they were before. Given the results of both T1 and T2, it seems reasonable to suggest that empathic concern does have a positive influence on a leader's ability to perform well in all three roles of responsible leadership simultaneously.

4.6. Testosterone

Male participants of our research study had the option of giving a saliva sample, which we then used to analyse their respective levels of the hormone testosterone. In total, 52 participants provided us with such a sample, one of which I

had to discard due to missing the corresponding activity log. While I hypothesised in Part 2.1.2 that testosterone would have an effect on responsible leadership effectiveness, I opted not to include it in my main regression model for two reasons. Firstly, research indicates that prenatal and current testosterone values are correlated (Sapienza et al. (2009)). Since I already measure the former with the right-hand 2D:4D, it seemed reasonable to leave out the latter due to potential multicollinearity issues. Secondly, the fact that I have far fewer data points for testosterone than for any other independent variable means there is a potential danger of overfitting. As mentioned in Subchapter 3.2, a common rule of thumb in regression analysis is that one ought to have a minimum of ten observed events for every independent variable. If I were to run a multiple regression with testosterone and eight or nine other independent variables, I would not be able to fulfil this condition.

Nevertheless, I was interested to see whether testosterone levels would affect any of the five responsible leadership dimensions. Therefore, I created a separate regression model which only includes the independent variable testosterone as well as the controlled variable age. There is no reason to control for gender in this model because all saliva samples came from men. I then ran this regression separately for each of the five dependent variables professional, facilitator, citizen, idea-provider, and total. The Durbin-Watson statistic for every model indicates that there is no issue with autocorrelation (the values are 1.990, 1.884, 2.177, 2.480, and 2.213 respectively). The error terms appear to be normally distributed, except in the model for idea-provider, which exhibits a slight positive skewness. I therefore performed a White test for this model ($p = 0.986$) and a Breusch-Pagan test for the other four ($p = 0.565, 0.493, 0.831, \text{ and } 0.417$). None of the five tests show any sign of heteroscedasticity. The VIFs of the coefficients further suggest that no multicollinearity is present. Table 27 on the next page shows the values for the independent variable testosterone in each of the five tested regression models. The full tables can be found in Appendix A.

As shown in the table, none of the models yielded a statistically significant result. Even the lowest p value, which is 0.217 for the model professional, is still far higher than the alpha of 0.05. I therefore cannot reject the null hypothesis that testosterone has no effect on the five dimensions of responsible leadership.

In addition to testosterone, we also analysed the levels of cortisol present in the saliva samples of our participants because recent research suggests that this hormone may have a moderating effect on the relationship between testosterone and aggressive behaviour. Popma et al. (2007) found that subjects with higher levels of testosterone were more likely

Table 24: Individual coefficients of T1

	Coefficient b	Std. Error	Beta	p value	VIF
(Constant)	2.944	3.747		0.434	
Gender	0.162	0.257	0.075	0.531	1.331
Age	-0.019	0.027	-0.078	0.475	1.090
Psychology	-0.059	0.246	-0.029	0.810	1.344
Honesty	0.062	0.197	0.034	0.755	1.115
Openness	-0.304	0.176	-0.184	0.087	1.055
Empathic	0.365	0.166	0.247	0.031	1.175
Perspective	-0.101	0.187	-0.063	0.590	1.246
Holism	0.113	0.315	0.040	0.722	1.186
2D:4D	-0.109	3.451	-0.003	0.975	1.098

Table 25: Summarised results for T2

R	R ²	Adj. R ²	SEE	F value	p value
0.278	0.078	0.057	0.993	3.824	0.025

Table 26: Individual coefficients of T2

	Coefficient b	Std. Error	Beta	p value	VIF
(Constant)	2.812	0.784		0.001	
Openness	-0.309	0.167	-0.187	0.068	1.011
Empathic	0.335	0.149	0.227	0.028	1.011

Table 27: Values for the variable testosterone in each of the five models

	Coefficient b	Std. Error	Beta	p value
Professional	-0.005	0.004	-0.177	0.217
Facilitator	0.004	0.003	0.171	0.233
Citizen	-0.001	0.003	-0.034	0.814
Idea-provider	0.003	0.003	0.140	0.331
Total	-0.003	0.004	-0.125	0.373

to display overt aggressive behaviour, but only if their cortisol levels were low. If both testosterone and cortisol levels were high, this effect would disappear. Since my hypothesis argues that testosterone influences responsible leadership effectiveness at least partly because it is related to aggressive behaviour, it seems reasonable to suggest that I would have to look at the ratio between testosterone and cortisol in order to obtain meaningful results. I therefore constructed five additional regression models, one for each of the dependent variables, in which I keep the independent variable age, but replace testosterone with C:T, the ratio between a participant's cortisol and testosterone levels.

Once again, the number of observations for every model is 51. The Durbin-Watson statistics are all relatively close

to two (1.985, 1.864, 2.181, 2.565, and 2.214), and the error terms for professional, facilitator, citizen, and total are approximately normally distributed, while idea-provider exhibits a slight positive skewness. The White test ($p = 0.692$) and the four Breusch-Pagan tests ($p = 0.744, 0.133, 0.678,$ and 0.524) give no indication of heteroscedasticity. The VIFs show no sign of multicollinearity either. Table 28 on the next page shows the results for the variable C:T. The full tables can be found in Appendix A.

As before, none of the models produce any statistically significant results for an alpha of 0.05. Thus, even when controlling for the potential influence of cortisol, I cannot find any support for the hypothesis that testosterone has an influence on responsible leadership effectiveness. Lastly, I

Table 28: Values for the variable C:T in each of the five models

	Coefficient b	Std. Error	Beta	p value
Professional	-0.101	0.235	-0.062	0.669
Facilitator	-0.208	0.165	-0.178	0.213
Citizen	-0.012	0.195	-0.009	0.950
Idea-provider	-0.229	0.155	-0.208	0.147
Total	-0.133	0.211	-0.088	0.531

looked at the correlation between the variables testosterone and 2D:4D. I obtained an r of 0.268 with a p value of 0.057. As such, the correlation is just barely nonsignificant for an alpha of 0.05.

5. Discussion

In Chapter 4, I presented the results of my empirical regression analysis without any value judgements. This chapter will now focus on the interpretation of said results. I will see how well they compare to my proposed hypotheses and attempt to find reasonable explanations for any discrepancies.

My analysis will follow the same basic structure as Subchapter 2.2 did. Thus, Subchapter 5.1 is concerned with the demographic and biological factors I measured, while Subchapter 5.2 addresses the personality traits. Additionally, Subchapter 5.3 will note more general observations about the models I have constructed, which are not related solely to one single variable. Lastly, Subchapter 5.4 will summarise the conclusions of my thesis and discuss avenues that future research could focus on.

5.1. Demographic and biological characteristics

This subchapter discusses the five demographic and biological traits that were measured in my regression models: gender, age, psychology student, and the 2D:4D as well as testosterone levels.

Gender

In Hypothesis 1, I stated that men are more likely to perform well in the role of professional because stereotypically male traits would aid in maximising the shareholder value, while women are more likely to perform well in the roles of facilitator and citizen because stereotypically female traits would be more conducive to interacting with employees and the external environment. Looking at the results from Chapter 4 reveals that this hypothesis is not supported by the data. Gender has no significant influence on any of the three responsible leadership roles.

The lack of a direct causal relationship may be due to a variety of reasons. As has been mentioned before, research studies focusing on gender differences in leadership have produced mixed results, and observed variations frequently disappear when looking at a formal organisational setting (Barbuto et al. (2007)). While our research study did not take place in such a setting, it is still conceivable that

by assuming the role of a leader in the think-aloud protocol, the participants adapted a different leadership approach than they would have used in a less standardised environment. Furthermore, Barbuto et al. (2007) also found that gender differences seem to be more apparent at lower levels of education, while disappearing at higher levels. Since our study focused almost exclusively on students from the UZH and the ETHZ, the gender differences would therefore be less pronounced.

Finally, it has to be pointed out that our research study did not assess the direct interactions between leaders and their relevant stakeholder groups; instead, we measured how strongly a participant considered the needs of his or her stakeholders. While it is possible that stereotypically female traits may indeed be useful for cooperating with employees and external associates, this does not necessarily mean that said traits would also make one more willing to actually consider the needs of these groups. Likewise, even if stereotypically male traits provide an advantage when it comes to generating shareholder value, there is no guarantee that these same traits would affect a person's desire to actually prioritise shareholder demands. Thus, the fact that I found no gender effect on the three responsible leadership roles does not necessarily mean that no such effect exists, but could rather indicate that our study design is inadequate for measuring it.

While I did not detect any significant influences for professional, facilitator, and citizen, gender differences do seem to have an impact on the role of idea-provider. Women scored noticeably higher on this dimension than their male peers. One potential explanation for this phenomenon is that it is caused by gender differences in creativity. Women may be more creative on average than men, which would allow them to come up with more inventive solutions and subsequently perform better as idea-providers. However, while some research on the subject does indicate that women perform better in specific domains of creative thinking (Averill (1999); Bowers (1971); Dudek et al. (1993)), the overall results are less conclusive. As Baer and Kaufman (2008) put it in their review of the literature: "In some cases ... there are significant numbers of studies in which one group or the other scores higher, but these are generally counter-balanced by studies showing just the opposite" (p. 28). These findings suggest that creativity on its own may not be sufficient to fully explain the observed gender differences.

Additional factors that might merit consideration are the potential differences in motivation or emotional involvement between genders. While a high level of creativity can certainly be advantageous for coming up with unique solutions, it in itself is meaningless if the person in question has no desire to actually think about alternative strategies. Thus, a person who performs well in the role of idea-provider would need both the means of thinking outside of the proverbial box and the inclination to actually do so. Within the confines of our research study, this could mean that participants who scored high on idea-provider were more willing to bring in their own thoughts and ideas because they felt more emotionally involved in the think-aloud protocol. Various studies have shown that women display more empathy than men on average (Gault and Sabini (2000); Toussaint and Webb (2005)). This could indicate that women experience a stronger connection to the plight of the various stakeholders involved in the think-aloud protocol, which in turn could cause them to come up with new solutions in an attempt to make everyone content. Meanwhile, men might have felt more distant from the presented problem on average, causing them to stick to more conventional approaches because they did not care enough about satisfying everyone to explore alternative ideas.

Age

Hypothesis 2 stated that age would have a positive effect on all three roles of responsible leadership (professional, facilitator, and citizen) as well as on the role of idea-provider. I justified this statement by arguing that older leaders tend to be more team-oriented and can use their accumulated experience to find better solutions to existing problems. However, my regression models do not support the hypothesis, as age was found to be nonsignificant for every dependent variable. This may be partly because, as discussed in Part 2.2.1, the relationship between leadership behaviour and age is more complex than Hypothesis 2 would suggest. Younger leaders tend to be highly motivated and flexible (Korac-Kakabadse et al. (1998)), traits that could potentially counterbalance some of the advantages older leaders possess.

Perhaps more importantly, the age span we surveyed in our empirical research study was extremely narrow. Almost all of our participants were between 18 and 30 years old. Only three participants were older than 30, and only one of those was older than 40. This limits our study's ability to gauge age-related effects, as minor differences in years would presumably not affect leadership behaviour to an enormous extent. Furthermore, age-related effects are often caused by differences in experience. Older leaders do not differ from their younger peers purely because of the additional years they have accumulated, but rather because during those years, they were able to collect additional experience in their respective field or organisation. Since almost all of our participants were students, it seems reasonable to assume that their experiences would be relatively similar. Hence, it is not too surprising that I was unable to find any support for the assumption that age affects responsible lead-

ership effectiveness; not necessarily because such effects do not exist, but because our study design does not allow us to optimally measure them. Controlling for age in my multiple regression models was still a worthwhile endeavour, but if one wanted to specifically assess the impact of age on responsible leadership behaviour, one would have to conduct a study whose subjects exhibit greater differences in age and correspondingly a large variety of unique experiences. Perhaps it would even make sense to replace age with a new variable that measures years of experience instead.

Psychology student

Hypothesis 3 argued that students who major in psychology would be more likely to perform well in the roles of facilitator, citizen, and idea-provider than students from other fields because they tend to show greater empathy and a higher degree of creativity. However, the results of my regression models do not support this statement. The variable psychology student was found to have no significant impact on facilitator and citizen. Additionally, while it did influence idea-provider, the relationship was negative, which means psychology students were less likely to score high on this dimension than students who major in other fields.

The lack of a causal relationship between psychology students and the roles facilitator and citizen may indicate that these students are not necessarily more empathic than the other participants we observed. My assumption in this regard was based on a single research study conducted with students from universities in Serbia (Dimitrijević et al. (2011)). It is possible that these results cannot be generalised without any caveats. Alternatively, the results of my regression analysis could also be caused by the fact that my models already control for empathy to a certain extent by assessing the personality characteristic empathic concern, which does have a significant effect on the role of facilitator. Looking at the bivariate correlations reveals that psychology student and empathic concern are positively correlated ($r = 0.207$ with $p = 0.045$). Thus, the variable psychology student might provide no additional explanatory power because its potential influence on the dependent variables is already covered by empathic concern.

The negative correlation between psychology student and idea-provider is similarly notable, as it would seemingly contradict the notion that those students are more creative than others. However, there are other potential explanations as well. One is that the psychology students simply lacked the necessary framework to conceptualise unique approaches to the problem. Our think-aloud protocol asks them to imagine themselves as CEOs of a multinational company, a role which they presumably did not encounter during their normal studies. Whereas students with a background in economics would be more accustomed to this type of scenario and consequently also more able to come up with highly inventive concepts. The difference in score could therefore be explained as a matter of experience rather than creativity. One limitation to this interpretation is that only a small number of non-psychology students actually majored in an eco-

nomie field and thus have a study-related advantage, while the other participants would be just as unfamiliar with these kinds of scenarios as the psychology students are.

Another, perhaps more convincing, explanation for this phenomenon lies in the study design. As mentioned in Subchapter 3.1, psychology students were the only participants in our research study who received a tangible compensation for their efforts, as they were granted hourly credits. Hence, while the other participants joined out of interest in the subject matter, some of the psychology students may have participated solely because of the promised reward. In turn, this might have affected their motivation and the level of emotional engagement they had with our think-aloud protocol. It is conceivable that they did not care as deeply about the presented scenario as the other participants on average and therefore did not wish to expand any strong effort on thinking about alternative solutions either.

Second-to-fourth digit ratio and testosterone

Hypotheses 5 and 6 were concerned with the 2D:4D and the testosterone level respectively. I argued that a low 2D:4D and high levels of testosterone would be positively correlated with the role of professional and negatively correlated with the roles of facilitator and citizen. However, this did not turn out to be the case in my regression models. Neither of the two variables had a significant effect on any of the five responsible leadership dimensions.

As discussed in Part 2.2.1, several researchers argue that testosterone increases the desire for social dominance (Mazur and Booth (1998)). Following this line of argumentation, those with a low 2D:4D or high levels of testosterone would be more likely to choose strategies which increase their social standing. In a corporate setting, this could mean that they are more willing to fulfil the role of professional because the shareholders have the best means of rewarding them for their behaviour. However, in our research study, the participants were not rewarded for maximising profits. Focusing on the role of professional to the exclusion of facilitator and citizen therefore did not provide them with any social advantages. It is possible that the lack of a specific reward structure in our think-aloud protocol would cause them to behave differently than they would have in a real-life example, neutralising the effects which form the foundation of my proposed hypotheses. Additionally, there is also the issue of accurately measuring the 2D:4D. In our study, we asked participants to keep their hand still and then directly assessed the length of the second and fourth digit using a Vernier caliper. While we measured both fingers twice and took the average of the two values to ensure greater consistency, our method is still susceptible to slight variations. Many of the observed ratios were very close to each other, so even a small error in the reported value could have a significant impact on the relative comparisons. Thus, the lack of significance could at least partly be the result of imprecise measuring. This would also explain why I could find no significant correlation between the 2D:4D and current testosterone levels. If further research studies wish to incorporate the 2D:4D in

their analysis, I would recommend relying on more precise measurement methods, for instance analysing photocopies of the respective hands.

5.2. Personality characteristics

This subchapter discusses the five personality characteristics I have observed: honesty-humility, openness to experience, empathic concern, perspective taking, and holism. As before, I will briefly restate my original hypothesis, compare it to the actual results from Chapter 4, and attempt to explain potential divergences.

Honesty-humility

In Hypothesis 6, I argued that those who score high on honesty-humility would perform better in the roles of facilitator and citizen because they are more willing to engage stakeholder demands even if there is no immediate financial incentive involved in doing so. However, the findings in Chapter 4 do not support my statement, as honesty-humility was not found to have a significant influence on any of the five dimensions of responsible leadership effectiveness. Though, as I have pointed out in Subchapter 3.2, Cronbach's alpha for honesty-humility was below the commonly accepted threshold of 0.7, indicating that the results for this variable may be less reliable than for the other dimensions of personality.

The lack of significant relationships could potentially indicate that honesty-humility is not as important to the responsible leadership process as the underlying theory would suggest, or it could mean that its impactful elements are already covered by different variables in the models. I have mentioned in Part 2.2.2 that honesty-humility is seen as an expression of fairness, but fairness is also commonly attributed to the personality dimension empathy (Page and Nowak (2002)). It is thus imaginable that adding honesty-humility to the model provides no additional explanatory power because its positive effects are already incorporated into empathic concern.

Additionally, the non-significance could also partly be the result of our study's design, specifically its lack of direct financial compensation for any of the participants. An important facet of honesty-humility is the dimension greed avoidance, which measures how strongly an individual is motivated by material gains. Those who score high on greed avoidance and correspondingly also high on honesty-humility overall tend to be less inclined to profit from the exploitation of others (Ashton and Lee (2007)). This behavioural trait may potentially encourage CEOs to act in a responsible manner in a real-life setting, as it would deter them from focusing on profit maximisation at the expense of everything else, but it does not play a large role in our think-aloud protocol because the individuals received no specific compensation regardless of whatever strategy they chose to pursue.

Openness to experience

Hypothesis 7 was concerned with the personality dimension openness to experience. Specifically, I argued that openness to experience would have a positive effect on a leader's ability

to perform well in the role of idea-provider due to its correlation with creativity. However, my observed data does not support this assumption, as I could find no significant relationship between the variables idea-provider and openness to experience. Somewhat surprisingly, openness to experience does have a significant impact on the role facilitator, but this relationship is negative. In other words, the higher an individual scored on openness to experience, the less likely he or she was to consider the needs of the employees.

The fact that openness to experience is nonsignificant for the role of idea-provider seems surprising, as this personality dimension is frequently considered to be strongly correlated with creative behaviour (Baer and Oldham (2006); McCrae (1987); Silvia et al. (2014)). Thus, one would assume that it simultaneously affects an individual's ability to come up with new and improved solutions to existing problems. However, as I have discussed before, our assessment of the role of idea-provider does not solely measure creativity, but also the motivation to actually employ it. It is possible that participants who score high on openness to experience are indeed more creative on average than their peers, but that does not necessarily mean that they were more interested in the presented scenario. If they did not care enough about the think-aloud protocol, they may simply have been unwilling to expend any effort on coming up with innovative approaches.

Perhaps the most surprising finding of my regression analysis is that openness to experience appears to have a negative influence on the role of facilitator. This relationship was found in both models that included openness to experience as a factor, F1 and F3, and has a p value that is considerably lower than the significance level of 0.05. The beta was -0.246 in F1 and -0.227 in F3, indicating a moderate influence. What makes these results so surprising is that research on the interactions between openness to experience and empathy has found a positive relationship; that is to say, higher levels of openness to experience correspond to higher levels of empathy (Magalhães et al. (2012)). Based on these results, one would expect that openness to experience also has a positive effect on facilitator, if any, but the opposite can be observed in my models.

It is difficult to pinpoint the exact cause of this phenomenon. A potential explanation may be the correlation between openness to experience and narcissism. Several studies have found that these two personality dimensions are related to each other (Furnham et al. (2013); Lee and Ashton (2005)); subjects who score high on openness to experience are also more likely to display narcissistic traits. Since narcissism is characterised by feelings of superiority, a desire for dominance, and a general willingness to exploit others for personal gain (Raskin and Terry (1988)), it could conceivably diminish one's ability to perform well in the role of facilitator.

On the other hand, the negative relationship could be a result of the specific scenario presented in our think-aloud protocol. It seems reasonable to assume that participants with a high openness to experience would also tend to be more open towards newer technologies. They may therefore

be more likely to implement the improved crude oil extraction method than others even if this decision harms the employees. If this interpretation is true, the negative impact of openness to experience on facilitator would be specific to our scenario, rather than being a general principle, and therefore not necessarily appear in further studies. However, both of the mentioned explanations are highly speculative in nature, and I would recommend that future research on the subject of responsible leadership continues to analyse the impact of openness to experience to see whether my findings can be confirmed or refuted.

Empathic concern

I argued in Hypothesis 8 that empathic concern would positively influence a leader's performance in the roles of facilitator and citizen because he or she would be more strongly affected by the plight of stakeholders and therefore also more motivated to find mutually beneficial solutions. The results of my empirical regression analysis partly support this hypothesis. Empathic concern does indeed seem to positively affect an individual's performance in the role of facilitator. However, it appears to have no significant effect on the role of citizen.

The positive relationship between empathic concern and facilitator is present in all three models F1 to F3 and significant even for a very low alpha of 0.001. The beta values are 0.407 for F1, 0.365 for F2, and 0.379 for F3, indicating that the impact of empathic concern is quite considerable. Furthermore, empathic concern is the only personality trait that affects a person's ability to perform all three roles simultaneously, showing a significant influence on the variable total in models T1 and T2. These results seem to confirm my initial assumption that empathy plays an important role in fostering responsible leadership effectiveness. It appears that highly empathic individuals are more likely to consider the needs of their employees and thereby act more responsibly in general. This may be either because they are better able to figure out what concerns said employees, because they are more willing to go out of their way to accommodate them, or a combination of both of these factors.

It is a bit surprising that empathic concern does not affect the performance as citizen, but this may be a result of the abstract nature of said role. In our think-aloud protocol, the negative consequences for the employees are immediately apparent. There is a direct causal link between implementing the new technology and having to fire parts of the staff. Conversely, it is not entirely clear how the environmental pollution will affect other human beings. While it is likely to have some negative impact, there is a degree of abstraction to the decision that makes it difficult to directly relate to the plight of the affected stakeholders. Participants who score high on empathic concern may thus be more likely to feel the suffering of the employees and try to ease their burden, while they are less concerned with the problems of external stakeholders, which are not as immediately obvious to them.

Perspective taking

In Hypothesis 9, I argued that scoring high on the person-

ality dimension perspective taking increases one's likelihood of performing well in the role of idea-provider. This hypothesis seems to be supported by the results of my empirical regression analysis. Perspective taking does indeed have a significant and positive effect on the dependent variable idea-provider, while having no influence on any of the other four dimensions of responsible leadership effectiveness.

Perspective taking's relationship with the role of idea-provider is significant in all three models I1 to I3 for an alpha of 0.05, though its p values are slightly higher than the ones for the variables gender and psychology student. The beta coefficient is 0.235 for I1, 0.236 for I2, and 0.199 for I3, indicating that the effect is moderately strong. These results do seem to support the theoretical notion that, unlike empathic concern, perspective taking does not directly affect an individual's performance in the three roles of responsible leadership, but it does aid in coming up with innovative and mutually beneficial solutions to existing problems. Hence, I posit that it is a valuable trait for establishing overall responsible leadership effectiveness.

Holism

My final hypothesis, Hypothesis 10, stated that holism would positively influence a leader's performance in the roles facilitator, citizen, and idea-provider because holistic thinking may help in solving complex, interwoven issues and is correlated with a greater concern for the environment. My models for the dependent variable citizen offer some evidence for part of this hypothesis, as holism has a significant and positive effect there. However, it does not even come close to having a significant effect on facilitator and idea-provider.

The relationship between holism and the role of citizen is significant in both models C1 and C2. The beta coefficient for holism is 0.309 in C1 and 0.269 in C2, which would indicate a moderately strong positive influence on the role of citizen. As mentioned, those who score high on holism report a greater connectedness with nature (Leong et al. (2014)), which might cause them to be more concerned about the potential environmental pollution that the new technology would bring than those who subscribe to a more analytic view of the world. In a similar vein, holistic thinkers would presumably be more likely to regard the company they work for as part of an interconnected system, rather than a completely separate institution. As such, they might be more receptive to the idea that harming the environment would negatively affect the entire network and therefore also the corporation in the long-term, causing them to give greater consideration to their responsibilities as citizen. However, it should be pointed out that holism had more missing observations than any of the other variables. While I used the expectation maximisation algorithm to fill in these empty data points, I cannot rule out the possibility that the missing entries reduce the reliability of the results.

Holism does not significantly impact the role of facilitator. Perhaps this is an indication that holistic thinkers are less concerned with what they perceive to be relatively small-scale issues. Their attention is centred on the larger, network-

spanning implications of the strategic decisions they have to undertake, and they may therefore not be as concerned with the plight of the individual employees. Holism's lack of influence on the role of idea-provider might be due to its positive correlation with perspective taking, a variable which is significant in all three models I1 to I3. It is possible that holism does not impact the role of idea-provider because its potential effects on creativity are already covered by perspective taking. In conclusion, while holism does not affect facilitator and idea-provider, my analysis does seem to suggest that it has a positive influence on citizen. As such, I would welcome it if further studies could take a more in-depth look at this potential relationship.

5.3. General observations

In addition to the individual personality characteristics that have been discussed in the previous two subchapters, there are a few general observations on my multiple linear regression models that are worth mentioning as well. For one, all constructed models have very low R^2 values. Even F1, which seems to be the most accurate model of the entire group, only has an R^2 of 0.213, indicating that almost 80% of the variation in the dependent variable facilitator cannot be explained by my chosen predictors. The adjusted R^2 is similarly low in all models and even negative in some select cases. However, these results were not entirely unexpected. Leadership is a multi-faceted phenomenon with a nearly unlimited number of potential influence factors. As such, even complex models will have trouble representing all relevant aspects. Furthermore, our study focused by design on the relationship between personality traits and responsible leadership effectiveness. We did not analyse the impact of behavioural attributes or contingency factors, which are considered to be of great importance in the theoretical literature (Derue et al. (2011); House and Aditya (1997)). It would therefore have been rather surprising if my models were able to explain all or even the vast majority of the variation in my dependent variables. My findings suggest that although traits play an important role in generating leadership effectiveness, they are not the only factors capable of doing so.

When looking at the individual models, it appears that the role of professional is the hardest one to properly assess and measure. In both P1 and P2, none of the observed independent variables are even remotely significant for an alpha of 0.05. Correspondingly, the p value for the F-test in both cases is extremely high. Neither P1 nor P2 thus seems to hold any explanatory power. This may simply indicate that my selection of traits was misguided. Perhaps the nature of the professional role is different from how I conceptualised it theoretically. I primarily argued that assertiveness and a desire for increased social status and dominance would cause leaders to focus more strongly on fulfilling shareholder needs and expectations, but there may be alternative aspects which I have failed to consider. It is conceivable that the role of professional is more heavily influenced by traits such as loyalty or devotion to the company. Furthermore, some participants in our study argued that the long-term survival of the firm

also benefits the employees who otherwise would lose their jobs. This line of thought suggests that the interactions between professional and facilitator are not as conflicting as some of my hypotheses indicate and that traits which positively influence one role would potentially also help in the other.

Another possible explanation for the lack of any significant variables in regard to the role of professional may be that the participants were already predisposed to give it special attention. While public discussions on leadership have put a stronger focus on ethical decision making in recent years, society still tends to view CEOs as being obligated towards their shareholders first and foremost (Waldman and Galvin (2008)). One would expect that the participants of our study reflect these values to some extent as well. In other words, it is possible that most participants, regardless of their respective personality traits, felt responsible for ensuring the survival of the corporation and the fulfilment of shareholder demands because they consider these tasks to be the primary duty of their position. This would also explain why the average score in the category professional was higher than for the other two roles (2.894 compared to 2.011 for facilitator and 2.872 for citizen). Furthermore, listening to the individual recordings of the think-aloud protocol reveals that while some participants did not analyse the needs of the employees or the external stakeholders in their answers, almost everyone gave at least some consideration to the needs of the company. This observation is congruent with the notion that professional is somehow seen as being more important or more pertinent to a CEO's activities than facilitator and citizen.

In contrast to professional, the models for facilitator were by far the most significant, with very low p values for the F-tests and the highest R^2 and adjusted R^2 . Looking at the individual results of the participants also reveals that the average score in the role of facilitator was lower than for the roles professional and citizen. This could indicate that the needs of the employees are treated as being less important, perhaps once again reflecting societal values. As discussed before, professional is indisputably seen as a central aspect of leadership. Meanwhile, recent corporate scandals, such as the BP oil spill or Nestlé's usage of unsustainable palm oil, have put a strong emphasis on the environmental components. Participants may thus have been primed to critically analyse any potential harm to the environment, whereas issues facing the workers are not as deeply embedded in their frame of reference. However, the lower average score in the role of facilitator may also have resulted from the wording of our think-aloud protocol. Our described setup could be understood primarily as a conflict between higher profits and greater harm to the environment, with the potential dismissal of employees being an ancillary problem. If some participants conceptualised it this way, it would explain why their score for professional and citizen is higher than the one for facilitator.

The models for the variable total are difficult to comprehensively analyse, as I did not create any specific hypotheses for them. Furthermore, the inter-rater reliability between my

own scores and those of my colleagues are lower for this dimension than the others, which is not too surprising given that it is harder to objectively assess how well an individual performs in all three roles simultaneously than how well he or she performs in each role separately. What is interesting to note is that the only significant variable in models T1 and T2 is empathic concern, with openness to experience having the second-lowest p value. Both variables also influence facilitator, which would suggest that those who performed better than average in this role also performed well in all three roles simultaneously. Once again, this may be due to the average score for facilitator being lower than for the other two roles. It seems that most participants were relatively good at performing as professional and citizen, which is why their score for all three roles together strongly depends on whether or not they also performed well as facilitator.

Lastly, it should be mentioned that the role of idea-provider had the lowest average score of all five dimensions (1.968). This may be partly due to the way it was measured, but it could also indicate that most participants were hesitant to bring in their own ideas and solutions. Lack of motivation may be a reason for this behaviour, although the fact that participation was voluntary and that many students received no material compensation for their efforts would suggest that they had at least some interest in the subject matter. Another explanation is that the participants felt they lacked the necessary expertise to offer unique solutions, as most of them were not enrolled in a major which is concerned with decision making in a business context. Finally, it could indicate that it did not even occur to them to try unconventional ideas, instead believing that relying on standard approaches would be the best idea. If this was the case, then perhaps fostering an individual's willingness to think outside of the box and move beyond conventional narratives would be an effective way of increasing his or her effectiveness as a responsible leader.

5.4. Conclusions

The goal of this master's thesis was to empirically assess and subsequently analyse the relationship between a variety of personality characteristics and responsible leadership effectiveness. For this purpose, I first established a set of ten demographic, biological, and personality factors and then hypothesised about their potential influence on responsible leadership behaviour. Using data gathered in an experimental research study in collaboration with members of the Chair of Clinical Psychology and Psychotherapy and the Chair of Foundations of Business Administration and Theories of the Firm at the UZH, I constructed a number of multiple linear regression models to test whether my assumptions would match reality. In this subchapter, I will summarise the most important results, briefly discuss the limitations of our study, and highlight aspects that future research could focus on. Table 29 on the next page lists all variables which were shown to be significant in one or more models for a significance level of 0.05.

Table 29: Values for the variable C:T in each of the five models

	Model	Coefficient b	Beta	p value
Gender	I1	-0.519	-0.275	0.018
	I3	-0.490	-0.260	0.019
Psychology student	I1	-0.556	-0.311	0.008
	I2	-0.397	-0.221	0.037
	I3	-0.602	-0.336	0.003
Openness to experience	F1	-0.362	-0.246	0.015
	F3	-0.333	-0.227	0.020
Empathic concern	F1	0.534	0.407	0.000
	F2	0.480	0.365	0.001
	F3	0.498	0.379	0.000
	T1	0.365	0.247	0.031
	T2	0.335	0.227	0.028
Perspective taking	I1	0.333	0.235	0.036
	I2	0.335	0.236	0.036
	I3	0.282	0.199	0.047
Holism	C1	0.793	0.309	0.007
	C2	0.689	0.269	0.012

Of my ten hypotheses, only one was fully supported by the empirical results: Hypothesis 9. Perspective taking is positively correlated with a leader's ability to perform well in the role of idea-provider in my regression models, suggesting that the ability to put oneself in someone else's shoes and see a situation from his or her point of view is a valuable trait to have as a responsible leader.

Additionally, I found partial support for Hypothesis 8, as empathic concern has a positive influence on the roles of facilitator and total. Those who score high on this personality dimension are thus more likely to consider the needs of the employees and the needs of all stakeholder groups simultaneously, which would indicate that responsible leadership effectiveness increases when someone is strongly concerned with the well-being of others. This finding is congruent with the theoretical literature in the field, which frequently ascribes an ethical dimension to responsible leadership behaviour (Maak and Pless (2006); Waldman (2011)). However, contrary to my expectations, empathic concern does not affect the role of citizen.

My results partially support Hypothesis 10, as holism has a significant impact on the role of citizen for an alpha of 0.05. This positive relationship is congruent with the theoretical literature and earlier research on the subject, which is why I believe that following up on this result would be of great value for the field of responsible leadership as a whole. However, it appears that holism has no influence on the roles of facilitator and idea-provider, despite my initial assumptions to the

contrary.

I also found some unexpected results that were not covered by my ten initial hypotheses. Gender and psychology student both had an influence on the role of idea-provider, as men and psychology students were less likely to perform well in this dimension than the other participants. With regard to the former, this result may be related to gender differences in empathy and therefore in emotional involvement in the think-aloud protocol. As for psychology students, the result may very well have been caused by the peculiarities of our study design and its reward structure.

Furthermore, openness to experience was shown to have a negative influence on the role of facilitator, which I did not foresee in Hypothesis 7. This may be because I have neglected its correlation with narcissism, or it may be a result of the specific subject matter of our think-aloud protocol. Either way, the result is surprising, and it would be interesting to see whether it can be replicated in subsequent studies.

None of the other hypotheses I have created were supported by the results of my regression analysis. Of the ten observed traits, four had no influence on any of the five responsible leadership dimensions: age, the 2D:4D, testosterone levels, and honesty-humility. To a certain extent, this is not unexpected. The lack of empirical studies on the subject meant that I did not have a wealth of data to go on when choosing the traits for my models. As such, I based my selection on theoretical notions or studies that have been undertaken in related fields. The non-significance of some

of these selected traits may indicate that my respective hypotheses were founded on faulty premises or that there exist additional aspects to these dimensions which I have failed to appropriately consider. At the same time, I cannot entirely exclude the possibility that these results were at least partly caused by limitations in our research study and that the non-significant variables may exhibit a significant influence in a study with a different design.

In general, I found that participants scored higher on average in the roles of professional and citizen than in the role of facilitator, perhaps reflecting societal values and expectations. Fostering their awareness for the needs of the employees would therefore seem to be a promising way to increase their overall responsible leadership effectiveness. Similarly, most participants scored low in the role of idea-provider, indicating that they were unwilling or unable to come up with innovative approaches. As with facilitator, training or educating them in this regard could be a valuable step in turning them into responsible leaders.

There are a number of limitations that have to be mentioned when critically evaluating our empirical research study. For one, we did not include any behavioural or contingency factors. Nor did we analyse the importance the organisational environment or previous work and leadership experiences play in creating responsible leadership effectiveness. These omissions were intentional, as we were solely focused on analysing the relationship between personality traits and responsible leadership, but they can be seen as interesting starting points for future research.

Since our input factors were collected through a self-assessment, there is the potential danger that the participants intentionally or unintentionally distorted the results by having an unrealistic view of themselves. Similarly, some of them may have given replies in the think-aloud protocol that they thought we wanted to hear, rather than answering completely truthfully. Furthermore, our think-aloud protocol is not necessarily a realistic representation of the actual leadership process. We did not incorporate a reward structure in our study, nor did we give participants the option to collect additional information or discuss the issue with stakeholders or other leaders. Lastly, our assessment of the responsible leadership roles was based on scales which have not been validated in previous empirical studies. As such, I cannot say with absolute certainty that they are a suitable measuring instrument.

In conclusion, I do find some support for the theoretical notion that personality traits affect responsible leadership effectiveness, although the results of my empirical regression analysis indicate that this interaction is perhaps less strong and far-reaching than I had initially assumed. My results suggest that empathic concern and openness to experience influence the role of facilitator and that holism influences the role of citizen. I believe that it is worth following up on these results and would therefore welcome it if further research included these three personality dimensions to see whether my results can be corroborated. Likewise, it would be valuable to see if the influence of gender, psychology stu-

dent, and perspective taking on the role of idea-provider can be replicated in subsequent studies, or if it is confined to my empirical regression analysis. Having additional research use our scales for measuring the five responsible leadership roles would also give us more opportunities to assess their validity and improve upon them if necessary.

In addition to building on top of my previously discussed results, it would also be great to see future research circumvent some of the limitations that inhibit the explanatory power of our empirical study. In particular, constructing a more realistic and in-depth representation of the leadership process with an actual reward structure could help to improve the validity of the obtained results and potentially give greater insights into aspects that have hitherto been neglected. In a similar vein, including situational and behavioural factors in addition to personality traits would grant a more comprehensive picture of responsible leadership, although the complexity of the subject and the huge number of potentially influential variables would make constructing such a study a challenging task.

In any case, I believe that undertaking further research on the relationship between personality traits and responsible leadership effectiveness is an important and worthwhile endeavour. There is little reason to assume that the degree of globalisation will decline in the near future. On the contrary, if past trends are any indication, the business environment of multinational companies will become even more complex and interwoven in the coming years. As such, I expect that responsible leadership will continue to play an important role and that understanding its antecedents will be of great value to both companies, who are looking for this type of leader to fill their top positions, as well as society as a whole, which benefits from having CEOs whose motivation does not solely lie in maximising profits and who understand the importance of considering stakeholders in their decision-making process.

References

- Adams, K. A. and Lawrence, E. K. *Research methods, statistics, and applications*. Sage Publications, 2014.
- Alin, A. Multicollinearity. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(3):370–374, 2010.
- Archer, J. Testosterone and human aggression: an evaluation of the challenge hypothesis. *Neuroscience & Biobehavioral Reviews*, 30(3):319–345, 2006.
- Ashton, M. C. and Lee, K. Empirical, theoretical, and practical advantages of the hexaco model of personality structure. *Personality and social psychology review*, 11(2):150–166, 2007.
- Ashton, M. C. and Lee, K. The hexaco–60: A short measure of the major dimensions of personality. *Journal of personality assessment*, 91(4):340–345, 2009.
- Ashton, M. C., Lee, K., Perugini, M., Szarota, P., De Vries, R. E., Di Blas, L., Boies, K., and De Raad, B. A six-factor structure of personality-descriptive adjectives: solutions from psycholexical studies in seven languages. *Journal of personality and social psychology*, 86(2):356–366, 2004.
- Averill, J. R. Individual differences in emotional creativity: Structure and correlates. *Journal of personality*, 67(2):331–371, 1999.
- Baer, J. and Kaufman, J. C. Gender differences in creativity. *The Journal of Creative Behavior*, 42(2):75–105, 2008.
- Baer, M. and Oldham, G. R. The curvilinear relation between experienced creative time pressure and creativity: moderating effects of openness to experience and support for creativity. *Journal of Applied Psychology*, 91(4):963–970, 2006.
- Baluch, B., Martin, G. N., Christian, L., and Corulla, W. J. Psychology and non-psychology students' estimation of their desirable and undesirable personality traits. *Personality and individual differences*, 21(4):617–620, 1996.
- Barbuto, J. E., Fritz, S. M., Matkin, G. S., and Marx, D. B. Effects of gender, education, and age upon leaders' use of influence tactics and full range leadership behaviors. *Sex Roles*, 56(1-2):71–83, 2007.
- Batson, D. Empathic concern and altruism in humans, 2009. URL <http://nationalhumanitiescenter.org/on-the-human/2009/10/empathic-concern-and-altruism-in-humans/>. Retrieved from.
- Beck, R. and Harter, J. Why good managers are so rare. *Harvard Business Review*, 2014.
- Bowers, K. S. Sex and susceptibility as moderator variables in the relationship of creativity and hypnotic susceptibility. *Journal of Abnormal Psychology*, 78(1):93–100, 1971.
- Breusch, T. S. and Pagan, A. R. A simple test for heteroscedasticity and random coefficient variation. *Econometrica: Journal of the Econometric Society*, pages 1287–1294, 1979.
- Bundy, J., Shropshire, C., and Buchholtz, A. K. Strategic cognition and issue salience: Toward an explanation of firm responsiveness to stakeholder concerns. *Academy of Management Review*, 38(3):352–376, 2013.
- Cameron, K. Responsible leadership as virtuous leadership. In *Responsible Leadership*, pages 25–35. Springer, 2011.
- Carli, L. L. and Eagly, A. H. Gender, hierarchy, and leadership: An introduction. *Journal of Social Issues*, 57(4):629–636, 2001.
- Carson, T. Friedman's theory of corporate social responsibility. *Business and Professional Ethics Journal*, 12(1):3–32, 1993.
- Choi, I., Koo, M., and Choi, J. A. Individual differences in analytic versus holistic thinking. *Personality and Social Psychology Bulletin*, 33(5):691–705, 2007.
- Coan, R. W. Measurable components of openness to experience. *Journal of Consulting and Clinical Psychology*, 39(2):346, 1972.
- Coates, J. M., Gurnell, M., and Rustichini, A. Second-to-fourth digit ratio predicts success among high-frequency financial traders. *Proceedings of the National Academy of Sciences*, 106:623–628, 2009.
- Dabbs Jr, J. M. and Morris, R. Testosterone, social class, and antisocial behavior in a sample of 4,462 men. *Psychological Science*, 1(3):209–211, 1990.
- Davis, M. H. Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of personality and social psychology*, 44(1):113–126, 1983.
- Day, D. V. and Antonakis, J. Leadership: Past, present, and future. *The nature of leadership*, 2:3–25, 2012.
- Derue, D. S., Nahrgang, J. D., Wellman, N., and Humphrey, S. E. Trait and behavioral theories of leadership: An integration and meta-analytic test of their relative validity. *Personnel psychology*, 64(1):7–52, 2011.
- Dimitrijević, A., Hanak, N., and Milojević, S. Psychological characteristics of future helping professionals: Empathy and attachment of psychology students. *Psihologija*, 44(2):97–115, 2011.
- Do, C. B. and Batzoglou, S. What is the expectation maximization algorithm? *Nature biotechnology*, 26(8):897–899, 2008.
- Dudek, S. Z., Strobel, M., and Runco, M. A. Cumulative and proximal influences on the social environment and children's creative potential. *The Journal of Genetic Psychology*, 154(4):487–499, 1993.
- Durbin, J. and Watson, G. S. Testing for serial correlation in least squares regression. ii. *Biometrika*, 38:159–178, 1951.
- Eagly, A. H. and Johnson, B. T. Gender and leadership style: A meta-analysis. *Psychological bulletin*, 108(2):233–256, 1990.
- Einolf, C. J. Empathic concern and prosocial behaviors: A test of experimental results using survey data. *Social Science Research*, 37(4):1267–1279, 2008.
- Eisenegger, C., Naef, M., Snozzi, R., Heinrichs, M., and Fehr, E. Prejudice and truth about the effect of testosterone on human bargaining behaviour. *Nature*, 463:356–359, 2010.
- Eisenegger, C., Haushofer, J., and Fehr, E. The role of testosterone in social interaction. *Trends in cognitive sciences*, 15(6):263–271, 2011.
- Elmudi, D., Minnis, W., and Abebe, M. Does education have a role in developing leadership skills? *Management Decision*, 43(7/8):1018–1031, 2005.
- Fernando, M. *Leading Responsibly in the Asian Century*. Springer, 2016.
- Figge, L. and Martens, P. Globalisation continues: The maastricht globalisation index revisited and updated. *Globalizations*, 11(6):875–893, 2014.
- Finkelstein, S. and Boyd, B. K. How much does the ceo matter? the role of managerial discretion in the setting of ceo compensation. *Academy of Management journal*, 41(2):179–199, 1998.
- Freeman, J. Towards a definition of holism. *Br J Gen Pract*, 55(511):154–155, 2005.
- Freeman, R. E. and Auster, E. R. Values, authenticity, and responsible leadership. In *Responsible Leadership*, pages 15–23. Springer, 2011.
- Furnham, A., Hughes, D. J., and Marshall, E. Creativity, ocd, narcissism and the big five. *Thinking Skills and Creativity*, 10:91–98, 2013.
- Galinsky, A. D. and Moskowitz, G. B. Perspective-taking: decreasing stereotype expression, stereotype accessibility, and in-group favoritism. *Journal of personality and social psychology*, 78(4):708–724, 2000.
- Galinsky, A. D., Maddux, W. W., Gilin, D., and White, J. B. Why it pays to get inside the head of your opponent: The differential effects of perspective taking and empathy in negotiations. *Psychological science*, 19(4):378–384, 2008.
- Gault, B. A. and Sabini, J. The roles of empathy, anger, and gender in predicting attitudes toward punitive, reparative, and preventative public policies. *Cognition & Emotion*, 14(4):495–520, 2000.
- George, J. M. and Zhou, J. When openness to experience and conscientiousness are related to creative behavior: an interactional approach. *Journal of applied psychology*, 86(3):513–524, 2001.
- Graziano, W. G. and Eisenberg, N. Agreeableness: A dimension of personality. In *Handbook of personality psychology*, pages 795–824. Elsevier, 1997.
- Graziano, W. G., Jensen-Campbell, L. A., and Hair, E. C. Perceiving interpersonal conflict and reacting to it: the case for agreeableness. *Journal of personality and social psychology*, 70(4):820–835, 1996.
- Graziano, W. G., Habashi, M. M., Sheese, B. E., and Tobin, R. M. Agreeableness, empathy, and helping: A person × situation perspective. *Journal of personality and social psychology*, 93(4):583–599, 2007.
- Harrell, F. E., Lee, K. L., and Mark, D. B. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in medicine*, 15(4):361–387, 1996.
- Harris, J. A. Measured intelligence, achievement, openness to experience, and creativity. *Personality and individual differences*, 36(4):913–929, 2004.
- Hines, M. Prenatal testosterone and gender-related behaviour. *European Journal of Endocrinology*, 155(suppl 1):115–121, 2006.
- Hoffman, B. J., Woehr, D. J., Maldagen-Youngjohn, R., and Lyons, B. D. Great man or great myth? a quantitative review of the relationship between individual differences and leader effectiveness. *Journal of Occupational and Organizational Psychology*, 84(2):347–381, 2011.

- House, R. J. and Aditya, R. N. The social scientific study of leadership: Quo vadis? *Journal of management*, 23(3):409–473, 1997.
- Howard, P. J. and Howard, J. M. The big five quickstart: An introduction to the five-factor model of personality for human resource professionals. *Charlotte, NC: Center for Applied Cognitive Studies*, 1995.
- Jago, A. G. Leadership: Perspectives in theory and research. *Management science*, 28(3):315–336, 1982.
- Jensen-Campbell, L. A. and Graziano, W. G. Agreeableness as a moderator of interpersonal conflict. *Journal of personality*, 69(2):323–362, 2001.
- Judge, T. A., Bono, J. E., Ilies, R., and Gerhardt, M. W. Personality and leadership: a qualitative and quantitative review. *Journal of applied psychology*, 87(4):765–780, 2002.
- Judge, T. A., Colbert, A. E., and Ilies, R. Intelligence and leadership: a quantitative review and test of theoretical propositions. *Journal of Applied Psychology*, 89(3):542–552, 2004.
- Kamm, F. M. Harming some to save others. *Philosophical Studies*, 57(3): 227–260, 1989.
- Korac-Kakabadse, A., Korac-Kakabadse, N., and Myers, A. Demographics and leadership philosophy: exploring gender differences. *Journal of Management Development*, 17(5):351–388, 1998.
- Laplume, A. O., Sonpar, K., and Litz, R. A. Stakeholder theory: Reviewing a theory that moves us. *Journal of management*, 34(6):1152–1189, 2008.
- Lechuga, J., Santos, B. M., Garza-Caballero, A. A., and Villarreal, R. Holistic reasoning on the other side of the world: Validation of the analysis-holism scale in mexicans. *Cultural Diversity and Ethnic Minority Psychology*, 17(3):325–330, 2011.
- Lee, K. and Ashton, M. C. Scale descriptions, n.d. URL <http://hexaco.org/scaledescriptions>. Retrieved from.
- Lee, K. and Ashton, M. C. Psychopathy, machiavellianism, and narcissism in the five-factor model and the hexaco model of personality structure. *Personality and Individual Differences*, 38(7):1571–1582, 2005.
- Leong, L. Y. C., Fischer, R., and McClure, J. Are nature lovers more innovative? the relationship between connectedness with nature and cognitive styles. *Journal of Environmental Psychology*, 40:57–63, 2014.
- LePine, J. A., Colquitt, J. A., and Erez, A. Adaptability to changing task contexts: Effects of general cognitive ability, conscientiousness, and openness to experience. *Personnel psychology*, 53(3):563–593, 2000.
- Little, R. J. A test of missing completely at random for multivariate data with missing values. *Journal of the American statistical Association*, 83(404): 1198–1202, 1988.
- Lowry, R. Kappa as a measure of concordance in categorical sorting, n.d. URL <http://vassarstats.net/kappa.html>. Retrieved from.
- Maak, T. Responsible leadership, stakeholder engagement, and the emergence of social capital. *Journal of Business Ethics*, 74(4):329–343, 2007.
- Maak, T. and Pless, N. M. Responsible leadership in a stakeholder society—a relational perspective. *Journal of business ethics*, 66(1):99–115, 2006.
- Maak, T., Pless, N. M., and Voegtlin, C. Business statesman or shareholder advocate? ceo responsible leadership styles and the micro-foundations of political csr. *Journal of Management Studies*, 53(3):463–493, 2016.
- Magalhães, E., Costa, P., and Costa, M. J. Empathy of medical students and personality: evidence from the five-factor model. *Medical teacher*, 34(10):807–812, 2012.
- Mandell, B. and Pherwani, S. Relationship between emotional intelligence and transformational leadership style: A gender comparison. *Journal of business and psychology*, 17(3):387–404, 2003.
- Mazur, A. and Booth, A. Testosterone and dominance in men. *Behavioral and brain sciences*, 21(3):353–363, 1998.
- McCormack, L. and Mellor, D. The role of personality in leadership: An application of the five-factor model in the Australian military. *Military Psychology*, 14(3):179–197, 2002.
- McCrae, R. R. Creativity, divergent thinking, and openness to experience. *Journal of personality and social psychology*, 52(6):1258–1265, 1987.
- McDonald, J. D. Measuring personality constructs: The advantages and disadvantages of self-reports, informant reports and behavioural assessments. *Enquire*, 1(1):1–19, 2008.
- McHugh, M. L. Interrater reliability: the kappa statistic. *Biochemia medica: Biochemia medica*, 22(3):276–282, 2012.
- Miska, C., Hilbe, C., and Mayer, S. Reconciling different views on responsible leadership: A rationality-based approach. *Journal of Business Ethics*, 125(2):349–360, 2014.
- Mitchell, R. K., Agle, B. R., and Wood, D. J. Toward a theory of stakeholder identification and salience: Defining the principle of who and what really counts. *Academy of management review*, 22(4):853–886, 1997.
- Neave, N., Laing, S., Fink, B., and Manning, J. T. Second to fourth digit ratio, testosterone and perceived male dominance. *Proceedings of the Royal Society of London B: Biological Sciences*, 270(1529):2167–2172, 2003.
- Nieziak, L. W., Siero, F. W., Dijkstra, P., Buunk, A. P., and Barelds, D. P. Empathic concern: Distinguishing between tenderness and sympathy. *Motivation and emotion*, 36(4):544–549, 2012.
- Oshagbemi, T. Age influences on the leadership styles and behaviour of managers. *Employee Relations*, 26(1):14–29, 2004.
- Oswald, P. A. The interactive effects of affective demeanor, cognitive processes, and perspective-taking focus on helping behavior. *The Journal of social psychology*, 142(1):120–132, 2002.
- Page, K. M. and Nowak, M. A. Empathy leads to fairness. *Bulletin of mathematical biology*, 64(6):1101–1116, 2002.
- Palazzo, G. and Scherer, A. G. Corporate legitimacy as deliberation: A communicative framework. *Journal of business ethics*, 66(1):71–88, 2006.
- Patzer, M. *Führung und ihre Verantwortung unter den Bedingungen der Globalisierung*. Berlin, Germany: Patzer Verlag, 2009.
- Paulus, C. Der Saarbrücker Persönlichkeitsfragebogen SPF (IRI) zur Messung von Empathie: Psychometrische Evaluation der deutschen Version des Interpersonal Reactivity Index, 2009. URL <http://psydok.sulb.uni-saarland.de/volltexte/2009/2363/>. Retrieved from.
- Pelletier, F. J. Holism and compositionality. *The Oxford handbook of compositionality*, pages 149–175, 2012.
- Perciavalle, V., Di Corrado, D., Petralia, M. C., Gurrisi, L., Massimino, S., and Coco, M. The second-to-fourth digit ratio correlates with aggressive behavior in professional soccer players. *Molecular medicine reports*, 7(6): 1733–1738, 2013.
- Pless, N. M. Understanding responsible leadership: Role identity and motivational drivers. *Journal of Business Ethics*, 74(4):437–456, 2007.
- Pless, N. M. and Maak, T. Responsible leadership: Pathways to the future. *Journal of Business Ethics*, 98:3–13, 2011.
- Popma, A., Vermeiren, R., Geluk, C. A., Rinne, T., van den Brink, W., Knol, D. L., Jansen, L. M., Van Engeland, H., and Doreleijers, T. A. Cortisol moderates the relationship between testosterone and aggression in delinquent male adolescents. *Biological psychiatry*, 61(3):405–411, 2007.
- Porter, M. E. and Kramer, M. Strategy and society: The link between competitive advantage and corporate social responsibility. *Harvard business review*, 84(12):78–92, 2006.
- Pringle, C. D., DuBose, P. B., and Yankey, M. D. Personality characteristics and choice of academic major: Are traditional stereotypes obsolete? *College Student Journal*, 44:131–142, 2010.
- Raskin, R. and Terry, H. A principal-components analysis of the narcissistic personality inventory and further evidence of its construct validity. *Journal of personality and social psychology*, 54(5):890–902, 1988.
- Ridgeway, C. L. Gender, status, and leadership. *Journal of Social issues*, 57(4):637–655, 2001.
- Russo, A. and Perrini, F. Investigating stakeholder theory and social capital: Csr in large firms and smes. *Journal of Business ethics*, 91(2):207–221, 2010.
- Sapienza, P., Zingales, L., and Maestripieri, D. Gender differences in financial risk aversion and career choices are affected by testosterone. *Proceedings of the National Academy of Sciences*, 106:15268–15273, 2009.
- Schroeder, D. A., Dovidio, J. F., Sibicky, M. E., Matthews, L. L., and Allen, J. L. Empathic concern and helping behavior: Egoism or altruism? *Journal of Experimental Social Psychology*, 24(4):333–353, 1988.
- Silvia, P. J., Beaty, R. E., Nusbaum, E. C., Eddington, K. M., Levin-Aspenson, H., and Kwapił, T. R. Everyday creativity in daily life: An experience-sampling study of “little c” creativity. *Psychology of Aesthetics, Creativity, and the Arts*, 8(2):183–188, 2014.
- Stanton, S. J., Liening, S. H., and Schultheiss, O. C. Testosterone is positively associated with risk taking in the Iowa gambling task. *Hormones and behavior*, 59(2):252–256, 2011.
- Stogdill, R. M. Personal factors associated with leadership: A survey of the literature. *The Journal of psychology*, 25(1):35–71, 1948.
- Tavakol, M. and Dennick, R. Making sense of cronbach's alpha. *International journal of medical education*, 2:53–55, 2011.
- Toussaint, L. and Webb, J. R. Gender differences in the relationship between empathy and forgiveness. *The Journal of social psychology*, 145(6):673–685, 2005.

- Van Bokhoven, I., Van Goozen, S. H., Van Engeland, H., Schaal, B., Arseneault, L., Séguin, J. R., Assaad, J.-M., Nagin, D. S., Vitaro, F., and Tremblay, R. E. Salivary testosterone and aggression, delinquency, and social dominance in a population-based longitudinal study of adolescent males. *Hormones and Behavior*, 50(1):118–125, 2006.
- VanVoorhis, C. W. and Morgan, B. L. Understanding power and rules of thumb for determining sample sizes. *Tutorials in Quantitative Methods for Psychology*, 3(2):43–50, 2007.
- Vedel, A. The big five and tertiary academic performance: A systematic review and meta-analysis. *Personality and Individual Differences*, 71:66–76, 2014.
- Viera, A. J., Garrett, J. M., et al. Understanding interobserver agreement: the kappa statistic. *Fam Med*, 37(5):360–363, 2005.
- Voegtlin, C. and Schwab, P. Good professional, colleague and citizen at the same time? an empirical examination of antecedents and outcomes of a role model of responsible leadership. *Unpublished manuscript, Department of Business Administration, University of Zurich, Switzerland*, 2015.
- Voegtlin, C. Development of a scale measuring discursive responsible leadership. In *Responsible Leadership*, pages 57–73. Springer, 2011.
- Voegtlin, C., Patzer, M., and Scherer, A. G. Responsible leadership in global business: A new approach to leadership and its multi-level outcomes. *Journal of Business Ethics*, 105(1):1–16, 2012.
- Vroom, V. H. and Jago, A. G. The role of the situation in leadership. *American psychologist*, 62(1):17–24, 2007.
- Waldman, D. A. Moving forward with the concept of responsible leadership: Three caveats to guide theory and research. *Journal of Business Ethics*, pages 75–83, 2011.
- Waldman, D. A. and Galvin, B. M. Alternative perspectives of responsible leadership. *Organizational Dynamics*, 37(4):327–341, 2008.
- Waldman, D. A. and Siegel, D. Defining the socially responsible leader. *The Leadership Quarterly*, 19(1):117–131, 2008.
- White, H. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: Journal of the Econometric Society*, pages 817–838, 1980.
- Winston, B. E. and Patterson, K. An integrative definition of leadership. *International journal of leadership studies*, 1(2):6–66, 2006.
- Yukl, G. A. Leadership in organizations (5th ed.). *Upper Saddle River, NJ: Prentice Hall*, 2002.
- Yukl, G. Managerial leadership: A review of theory and research. *Journal of management*, 15(2):251–289, 1989.
- Zaccaro, S. J. Trait-based perspectives of leadership. *American psychologist*, 62(1):6–16, 2007.
- Zaccaro, S. J., Kemp, C., and Bader, P. Leader traits and attributes. *The nature of leadership*, pages 101–124, 2004.



Zur experimentellen Analyse des Zusammenhangs von Feedback, Selbstwirksamkeit und Kreativität

Lea Katharina Haffke

Albert-Ludwigs-Universität Freiburg

Abstract

Kreativität – eine Eigenschaft, die im modernen Arbeitsumfeld als unabdingbar gilt. Trotz der hohen Nachfrage nach kreativen MitarbeiterInnen gibt es jedoch weiterhin keine universale Methode, wie Kreativität im Arbeitskontext wirksam gefördert werden kann. Feedback, als weit verbreitetes Führungsinstrument, ist ein kontextabhängiger Faktor, der Kreativität fördern oder hemmen kann. Das Ziel der vorliegenden Arbeit ist es, die Bedingungen unter denen Feedback einen Effekt auf die Kreativität von Personen haben kann, zu untersuchen. Des Weiteren wird betrachtet, welche Bedeutung das psychologische Konzept der Selbstwirksamkeitserwartung in diesem Wirkungszusammenhang hat. Es werden die theoretischen Zusammenhänge zwischen Feedback, Selbstwirksamkeit und Kreativität literaturbasiert aufgearbeitet, um diese anhand der Empirie aus ausgewählten experimentellen Studien zu überprüfen. Abschließend werden zwei eigenständig konzipierte experimentelle Designs vorgestellt, die bestehende Forschungslücken schließen. Die Ergebnisse zeigen, dass Feedback Informationen über die Anforderungen von kreativen Aufgaben vermittelt und daher einen positiven Effekt auf die Kreativität hat. Darüber hinaus steigert Feedback die intrinsische Aufgabenmotivation, welche als wichtiger Vorläufer von Kreativität angesehen wird. Diese kreativitätsfördernden Effekte sind jedoch nur unter der Bedingung möglich, dass Feedback informativ und aufgabenbezogen formuliert wird.

Keywords: Kreativität, Feedback, Selbstwirksamkeit, experimentelle Designs

1. Einleitung

„Unternehmen brauchen Kreativität als Kernressource.“¹; „Führungskräfte müssen bei kreativen Prozessen die Mitarbeiter führen wie ein Dirigent ein Orchester.“²; „Das Büro als kreative Keimzelle.“³. Die Zeitungartikel aus Deutschlandfunk und Zeit stehen exemplarisch für die Vielzahl der Artikel zum Thema Kreativität am Arbeitsplatz, welche die Aktualität dieses Themas darstellen. Jedoch lässt sich an den vielfältigen Beiträgen, von Ratschlägen über Kreativitätstechniken bis hin zu Thesen zu kreativen Persönlichkeitsfaktoren, eine Unsicherheit darüber ausmachen, mittels welcher Faktoren Kreativität im Arbeitskontext tatsächlich gefördert werden kann. Während sich die Kreativitätsforschung anfangs hauptsächlich mit den Eigenschaften kreativer Menschen befasst hat⁴, liegt der Fokus mittlerweile immer häufiger auf

den kontextabhängigen Faktoren zur Förderung und Hemmung von Kreativität.⁵ Feedback, das am häufigsten eingesetzte Werkzeug von Führungskräften, um Angestellte zu motivieren, zu leiten und zu informieren⁶ ist ein solcher Faktor. Feedback kann eine starke Auswirkung auf die Kreativität von Individuen haben.⁷ Jedoch wurde bisher wenig experimentelle Forschung zur Wirkung von Feedback auf Kreativität betrieben.⁸

Diese Arbeit leistet einen Beitrag zum Verständnis des Zusammenhangs von Feedback und Kreativität. Es wird sowohl der direkte Effekt betrachtet, als auch das psychologische Konzept der Selbstwirksamkeitserwartung, miteinbezogen, um die Mechanismen der Wirkung darzustellen. Ebenso werden Bedingungen für die Wirkung von Feedback auf Kreativität formuliert. Dementsprechend lässt sich folgende Forschungsfrage ableiten. Unter welchen Bedingungen hat Feed-

¹Quarch (2018).

²Dörflinger (2016)

³Quarch (2018).

⁴Vgl. Chan (2013), S. 24.

⁵Vgl. Shalley und Gilson (2004), S. 34.

⁶Vgl. Ashford und Cummings (1983), S. 371; Ilgen et al. (1979), S. 350.

⁷Vgl. Zhou (2008), S. 125.

⁸Vgl. Beckers et al. (2010), S. 476.

back einen Effekt auf die Kreativität von Personen?

Die theoretische Grundlage zur Beantwortung der Forschungsfrage wird geschaffen, indem im ersten Teil der Arbeit zunächst Kreativität, Feedback und Selbstwirksamkeit literaturbasiert erläutert werden. Auf dieser Grundlage werden die theoretischen Kausalzusammenhänge zwischen den Komponenten erklärt. Um die theoretischen Zusammenhänge anhand der Empirie zu prüfen, werden im zweiten Teil der Arbeit ausgewählte experimentelle Studien erörtert, verglichen und kritisch bewertet. Die im zweiten Teil ermittelten Forschungslücken und Limitationen der experimentellen Studien führen zu den zwei eigenständig konzipierten experimentellen Designs. Abschließend werden die Ergebnisse mit Bezug auf die Forschungsfrage resümiert.

2. Begriffsbestimmungen und theoretische Zusammenhänge

In dem folgenden Kapitel werden die grundlegenden Begriffe Kreativität, Feedback und Selbstwirksamkeit erläutert um eine theoretische Grundlage für die darauffolgende Analyse der Zusammenhänge zu leisten. In Kapitel 2.1 soll zuerst Kreativität bestimmt werden. Darauffolgend wird das Komponentenmodell der Kreativität vorgestellt, das die Determinanten für die individuelle kreative Leistung beschreibt. Des Weiteren, werden zwei Arten, Kreativität zu operationalisieren präsentiert. In Kapitel 2.2 wird die Definition des Begriffs Feedback, wie er in dieser Arbeit verwendet wird unter Betrachtung unterschiedlicher Definitionen herausgearbeitet. Im Anschluss wird die Wirkung von Feedback auf das Verhalten von Personen erläutert. In Kapitel 2.3 wird zuerst das psychologische Konzept der Selbstwirksamkeit in die sozial-kognitive Lerntheorie eingebettet, definiert und die Auswirkungen sowie die Einflussquellen dargelegt. Schlussendlich werden die literaturbasierten, theoretischen Zusammenhänge von Feedback, Selbstwirksamkeit und Kreativität in Kapitel 2.4 dargestellt.

2.1. Kreativität

In der psychologischen Forschung wird der Begriff Kreativität auf mehrere Aspekte, die „four P's of creativity“, bezogen.⁹ Diese umschließen person, process, product und press. Während bei person die Persönlichkeitsfaktoren im Vordergrund stehen, die zur Kreativität der Person beitragen, liegt bei process der Fokus auf den psychischen Mechanismen, die in einer kreativ denkenden Person vor sich gehen. Bei dem Aspekt product ist das kreative Endprodukt wesentlich, also beispielsweise ein Kunstwerk oder eine Erfindung. Zuletzt wird bei press der Einfluss der Umgebung auf das kreative Denken untersucht.¹⁰ Die Definition von Kreativität konzentriert sich meist ausschließlich auf den Aspekt product, da angenommen wird, dass ein kreatives Produkt in einem kreativen Prozess von einer kreativen Person in einem kreativen

Umfeld erschaffen worden sein muss.¹¹ Daran schließt Amabile (1983) konzeptuelle Definition an, die von der überwiegenden Mehrheit verwendet wird¹² und für die vorliegende Arbeit maßgeblich ist.

„A product or response will be judged as creative to the extent that (a) it is both a novel and appropriate, useful, correct or valuable response to the task at hand, and (b) the task is heuristic rather than algorithmic.“¹³

Es lassen sich drei Merkmale von kreativen Produkten bzw. Antworten festhalten. Erstens ist etwas nie per se kreativ, sondern wird von Personen als kreativ bewertet. Zweitens soll ein kreatives Produkt nicht nur neuartig, sondern auch nützlich sein, das heißt einen angemessenen Beitrag zur Lösung eines Problems liefern. Drittens sind laut Amabile nur bei heuristischen Aufgaben kreative Leistungen möglich.

Amabiles Prozesstheorie, das sogenannte Komponentenmodell der Kreativität, über die Natur und Bestandteile des kreativen Prozesses setzt an der oben genannten Definition an.¹⁴ Obwohl die Komponenten des kreativen Prozesses beschrieben werden, beschränkt sich das Modell nicht nur auf den process Aspekt, sondern umfasst ebenso individuelle (person) wie kontextuelle (press) Aspekte.¹⁵ Wegen dieser ganzheitlichen Sichtweise ist dieses Modell von Bedeutung für die vorliegende Arbeit, wo im weiteren Verlauf der individuelle Aspekt Selbstwirksamkeit und der kontextuelle Aspekt Feedback im Zentrum stehen. Kreativität wird also zum einen durch die individuellen Faktoren der Fähigkeiten und Kenntnisse eines Individuums sowie durch dessen Motivation bedingt. Und zum anderen durch den kontextuellen Faktor der sozialen Umwelt.¹⁶ Die Fähigkeiten beziehen sich auf den Bereich, in dem kreatives Verhalten beobachtet wird und umfassen berufsspezifisches Wissen, technische Fähigkeiten und besondere Talente in dem besagten Bereich.¹⁷ Amabile bezeichnet diese Fähigkeiten als „the set of cognitive pathways for solving a given problem or doing a given task“¹⁸. Um das entsprechende Problem oder die Aufgabe kreativ zu lösen, ist jedoch zusätzlich die Kenntnis kreativitätsrelevanter Prozesse vonnöten. Sie erweitert die bereichsrelevanten Fähigkeiten um einen förderlichen kognitiven Stil, Kenntnisse von Heuristiken zur Produktion kreativer Ideen (Kreativitätstechniken) und einen angemessenen Arbeitsstil. Ein kognitiver Stil förderlich zum kreativen Denken, sollte von Verständnis für Komplexität, Offenhalten von Antwortalternativen, Auf-

¹¹Vgl. Chan (2013), S. 27.

¹²Vgl. Baer et al. (2003), S. 570; Benzer (2006), S. 3; Byron und Khazanchi (2012), S. 810; Charness und Grieco (2014), S. 6; Fodor und Carver (2000), S. 383; Gross (2016), S. 1; Shalley (1991), S. 179; Wang et al. (2015), S. 189; Zhou (1998), S. 261.

¹³Amabile (1983b), S. 33.

¹⁴Vgl. Kozbelt et al. (2010), S. 30.

¹⁵Vgl. Amabile (1983b), S. 65.

¹⁶Vgl. Amabile (2013), S. 136.

¹⁷Vgl. Amabile (1983b), S. 67.

¹⁸Amabile (1983b), S. 69.

⁹Vgl. Kozbelt et al. (2010), S. 24.

¹⁰Vgl. Kozbelt et al. (2010), S. 24f.

schub von Bewertung und der Fähigkeit mentale Sets¹⁹ bei der Problemlösung zu überkommen, gekennzeichnet sein.²⁰ Daneben sind Heuristiken, das heißt individuelle Regeln, die Individuen bei der Herangehensweise an Probleme helfen, behilflich.²¹ Letztlich dient auch ein angemessener Arbeitsstil der Kreativität, wenn er durch langfristige Konzentrationsfähigkeit, sowie durch Beharrlichkeit bei Schwierigkeiten und durch ein hohes Level an Produktivität charakterisiert ist.²² Wie sich erkennen lässt, sind einige der oben benannten kreativitätsrelevanten Prozesse erlernbar, während andere von Persönlichkeitsfaktoren der Individuen abhängen.²³ Solche Persönlichkeitsfaktoren sind beispielsweise Selbstdisziplin, unabhängiges Urteilsvermögen, Non-Konformität, Risikofreude und Unabhängigkeit von gesellschaftlicher Akzeptanz.²⁴ Zusammengefasst, demonstrieren die berufsspezifischen Fähigkeiten und die kreativitätsspezifischen Prozesse die potentielle Leistungsfähigkeit eines Individuums im kreativen Prozess. Ob und wie diese potentiellen Fähigkeiten aber im kreativen Prozess eingesetzt werden, entscheidet die aufgabenspezifische Motivation.²⁵ Deren zentrale Bedeutung unterstreicht das intrinsic motivation principle. dessen Grundannahme lautet, dass die intrinsische Motivation kreatives Denken fördert, während die extrinsische Motivation in diesem Kontext schädlich ist.²⁶ Intrinsisch beschreibt hierbei eine Motivation, die aus der positiven Reaktion eines Individuums auf Qualitäten der Aufgabe selbst entsteht. Diese Reaktion zeigt sich durch Interesse, Engagement oder Neugier.²⁷ Extrinsisch motiviertes Verhalten hingegen wird nur um der folgenden Konsequenzen willen ausgeführt, also nicht aus Interesse an der Tätigkeit selbst.²⁸ Kreatives Denken ist demnach nur möglich, wenn das Individuum sich frei von Kontrolle entfalten und sich somit nur auf die Aufgabe als solche konzentrieren und ohne externen Druck arbeiten kann.²⁹ Die aufgabenspezifische Motivation wird neben der Einstellung zur Aufgabe auch von der Wahrnehmung der Gründe für die Durchführung der Aufgabe charakterisiert. Das erste Element wird dabei durch die kognitive Bewertung der Aufgabe und die Übereinstimmung mit bestehenden Präferenzen geformt. Das zweite Element, die Gründe für Durchführung der Aufgabe, ist weniger persönlich, sondern hauptsächlich von sozialen und umweltbezogenen Faktoren geprägt. Damit sind insbesondere die An- oder Abwesenheit äußerer Zwänge in der sozialen Umwelt gemeint. Äußere Zwänge entstehen durch die Präsenz kontrollierender, extrinsischer Fakto-

ren.³⁰ Die Anwesenheit dieser Zwänge verringert die intrinsische Aufgabenmotivation, indem der Fokus des Individuums nicht mehr auf der Lösung des Problems liegt, sondern auf der Erreichung des extrinsischen Ziels.³¹ Jedoch gibt es ebenfalls Ausnahmen von dem oben benannten intrinsic motivation principle. Wie Amabile (1996) in der Aktualisierung ihres Werks unter Bezug auf diverse empirische Studien einräumt, sind nur extrinsische Anreize schädlich, die vom Individuum als Leistungskontrolle empfunden werden.³² Anreize, die Informationen über die Anforderungen der Aufgabe liefern oder Kompetenzen bestätigen ohne Kontrolle auszuüben, sind für die Kreativität von Individuen nicht schädlich sondern sogar förderlich.³³ Die vierte Komponente, die außerhalb des Individuums wirkt, ist die soziale Umwelt.³⁴ Innerhalb dieser, gibt es einerseits Faktoren, die die intrinsische Motivation und damit Kreativität fördern und andererseits hemmen. Fördernde Faktoren sind laut Amabile (2013) kollaborative Arbeitsgruppen bestehend aus Personen mit vielfältigen Kompetenzen, Freiheit bei der Ausführung der Arbeit, ermutigende Vorgesetzte, eine globale kreativitätsfördernde Vision, sowie angemessene Anerkennung von kreativen Ideen. Kreativitätshemmende Faktoren bilden eine Atmosphäre, in der neue Ideen stark kritisiert werden, eine risikoaverse Stimmung in der Führungsspitze besteht, die Betonung auf dem Status Quo gelegt wird, anstatt eine Zukunftsvision aufzubauen und übermäßiger Zeitdruck herrscht.³⁵

Zusätzlich zur Beschreibung der Komponenten, setzt Amabile (1996) diese in Bezug zum fünfstufigen kreativen Prozess. Diese Stufen sind das Aufgreifen der Problemstellung, die Vorbereitung der Lösungssuche, die Entwicklung von Lösungsvorschlägen, die Bewertung der vorher entwickelten Lösungsvorschläge und schlussendlich das Ergebnis.³⁶ Abbildung 1 im Anhang bildet diesen Prozess grafisch ab. Zusätzlich wird erläutert welche Komponenten auf welchen Stufen wirken.

Kommt es zu der Frage wie Kreativität operationalisiert werden kann, gibt es seit Anbeginn der Kreativitätsforschung unterschiedliche Ansätze.³⁷ Für die vorliegende Arbeit ist insbesondere die subjektive Bewertungsmethode von Amabile (1983b) und der sogenannte Torrance Test of Creative Thinking (TTCT) relevant. Erstere wird als Consensual Assessment Technique (CAT) bezeichnet und fußt auf ihrer konsensualen Definition von Kreativität, welche die oben erwähnte konzeptuelle Definition um Informationen bezüglich des Bewertungsprozesses von kreativen Produkten ergänzt.

„A product or response is creative to the extent that appropriate observers independently agree it is creative. Appropriate observers are those familiar with the domain in which the product was

¹⁹Mentale Sets sind unterbewusste Tendenzen, Probleme in bestimmter Weise anzugehen, die häufig ein Hindernis bei der Lösungsfindung darstellen (Vgl. Bruno (2002), S. 127).

²⁰Vgl. Amabile (1983b), S. 72.

²¹Vgl. Amabile (1983b), S. 73.

²²Vgl. Amabile (1983b), S. 74.

²³Vgl. Amabile (1983b), S. 68.

²⁴Vgl. Amabile (1983b), S. 74.

²⁵Vgl. Diehl und Munkes (2002), S. 370.

²⁶Vgl. Amabile (1983b), S. 76.

²⁷Vgl. Amabile (1996), S. 115.

²⁸Vgl. Deci und Ryan (1985), S. 253f..

²⁹Vgl. Amabile (1983b), S. 94-96.

³⁰Vgl. Amabile (1983b), S. 76.

³¹Vgl. Crutchfield (1962), S. 125.

³²Vgl. Amabile (1996), S. 119.

³³Vgl. Amabile (1996), S. 117-119; Eisenberger et al. (1998), S. 710.

³⁴Vgl. Amabile (1996), S. 119-121.

³⁵Vgl. Amabile (2013), S. 136.

³⁶Vgl. Amabile (1996), S. 113.

³⁷Vgl. Cropley und Kaufman (2013), S. 196.

created or the response articulated. Thus, creativity can be regarded as the quality of products or responses judged to be creative by appropriate observers, and it can also be regarded as the process by which something so judged is produced.”³⁸

Diese Bewertungsmethode konzentriert sich auf die subjektive Bewertung durch RichterInnen. Grundannahme für das Vertrauen in die Qualität der subjektiven Bewertung, sind die Vermutung, dass Kreativität, auch wenn ihre Eigenschaften schwer zu beschreiben sind, von Personen einerseits erkannt wird und andererseits Einigkeit bei der Beurteilung herrscht.³⁹ Aufbauend auf diesen Annahmen, führt Amabile die Besonderheiten der Beurteilungstechnik aus, die sich vor allem auf die Vorgaben zu dem Bewertungsverfahren beziehen. Es werden unter anderem folgende Anforderungen an den Prozess gestellt. Erstens, gilt es angemessene Personen als RichterInnen auszuwählen, also solche, die bereits Erfahrungen in dem fraglichen Gebiet und dadurch implizite Kreativitätskriterien entwickelt haben. Zweitens, sollen die RichterInnen unabhängig voneinander und vor allem unabhängig von dem/der ExperimentatorIn die Kreativität der Problemlösungen bewerten, damit sie ihre impliziten Kriterien nutzen. Drittens, ist es essentiell den RichterInnen mitzuteilen, dass die Lösungen relativ zueinander bewertet werden sollen und nicht anhand eines absoluten Standards, den sie für den ihnen bekannten Bereich haben. Viertens, ist es ferner wichtig, die Lösungen in einer unterschiedlichen Reihenfolge zu zeigen, damit mögliche Reihenfolgeeffekte vermindert werden können. Eine Vorgabe für die Anzahl der RichterInnen gibt Amabile (1982) bewusst nicht, da nicht nur die Anzahl, sondern auch die durchschnittliche Korrelation zwischen den Bewertungen der RichterInnen zur Interrater-Reliabilität⁴⁰ beitragen. Demgemäß ist, abhängig von der Schwierigkeit der Bewertungsleistung, die Anzahl der RichterInnen zu wählen.⁴¹

Außerdem muss sich der/die ExperimentatorIn für ein Format der Bewertungsskala entscheiden. Mit Orientierung an der bestehenden Forschungsliteratur bietet sich eine 5- bis 6-Punkte Skala⁴² oder eine 10- bis 11- Punkte Skala⁴³ an. Sind die Daten erhoben, sind diese auf ihre Interrater-Reliabilität zu analysieren. Wenn dadurch die Verlässlichkeit der Bewertungen festgestellt wurde, lässt sich mit den Kreativitätswerten arbeiten.⁴⁴

Der TTCT fällt unter die Kategorie der Kreativitätstests, im Besonderen unter den Verhaltenstest. Diese von Torrance

(1974a) entwickelte Testbatterie umfasst sowohl verbale als auch figürliche Aufgaben und beruht auf einer prozessorientierten Definition von Kreativität als

„a process of becoming sensitive to problems, deficiencies, gaps in knowledge, missing elements, disharmonies, and so on; identifying the difficulty; searching for solutions, making guesses, or formulating hypotheses about the deficiencies; testing and retesting these hypotheses and possibly modifying and retesting them; and finally communicating the results.“⁴⁵

Im Einklang mit der bestehenden Forschungsliteratur⁴⁶ wird in der vorliegenden Arbeit nur eine Aufgabe und deren Wertung aus der Testbatterie beschrieben, die Unusual Uses Aufgabe. Diese ist eine Modifikation des Brick Uses Test von Guilford (1967a)⁴⁷, der im Rahmen seiner Intelligenzforschung zu den divergenten Fähigkeiten⁴⁸ entwickelt wurde. Bei dieser Aufgabe sollen möglichst ungewöhnliche Verwendungszwecke für einen Ziegelstein gefunden werden.⁴⁹ Torrance (1974a) ersetzt den Ziegelstein jedoch durch einen Pappkarton (Version A) und eine Konservendose (Version B). Die Aufgabe dient der Überprüfung, ob eine Testperson dazu fähig ist, seine/ihre mentalen Sets bei der Aufgabenlösung zu überwinden. Die kreativen Fähigkeiten werden dadurch demonstriert, dass sich die Testperson nicht auf eine Kategorie beschränken.⁵⁰ Die Testpersonen erzielen bei der Unusual Uses Aufgabe, ebenfalls mit Bezug auf Guilford (1967a), Punkte auf den drei Dimensionen fluency, flexibility und originality.⁵¹ Dabei bezeichnet fluency die Anzahl von verschiedenen Verwendungszwecken (und wird mit einem Punkt pro Antwort gewertet), bzw. allgemeiner die Anzahl der generierten kreativen Ideen. Flexibility bezieht sich auf die Anzahl der Kategorien, in die die Verwendungszwecke eingeteilt werden können. Je unterschiedliche Kategorie wird ein Punkt verteilt. Generalisierend zeigt diese Dimension, ob Testpersonen in der Lage sind, ihre Strategien zur Problemlösung während des Prozesses zu ändern. Originality bezeichnet die statistische Seltenheit im Vergleich zu retrospektiven Vergleichswerten. Universell wird damit die Fähigkeit abgebildet, nicht-offensichtliche und unkonventionelle Ideen zu generieren.⁵² Bei der Testbatterie wird abschließend für jede Testperson jeweils ein Score für Fluency, Flexibility und Originality errechnet. Da es sich bei der Anwendung der Unusual Uses Aufgabe aber nur um einen Einzeltest handelt und bei den meisten Studien ein vollständiger Kreativitätsscore von Interesse ist,

³⁸Amabile (1983b), S. 31.

³⁹Vgl. Barron und Harrington (1981), S. 5.

⁴⁰Die Interrater-Reliabilität, bzw. Urteilerübereinstimmung bezeichnet das Ausmaß, in dem zwei oder mehr unabhängige Beurteiler (Rater) in ihrem Einschätzungsergebnis übereinstimmen. (Vgl. Wirtz (2014), S. 276).

⁴¹Vgl. Amabile (1982), S. 1002f.

⁴²Vgl. u. a. Batey et al. (2010), S. 533; Friedman (2009), S. 260; Pretz und McCollum (2014), S. 230.

⁴³Vgl. Benzer (2006), S. 19; Zhou (1998), S. 268.

⁴⁴Vgl. Amabile (1982), S. 1002.

⁴⁵Torrance (1974b), S. 8.

⁴⁶Vgl. u. a. Batey et al. (2010), S. 533; Bradler et al. (2016), S. 6; Furnham et al. (2008), S. 1064.

⁴⁷Vgl. Torrance (1974b), S. 12.

⁴⁸Divergentes Denken bezieht sich auf den Denkprozess, bei dem in viele verschiedene Richtungen gedacht und das Ziel keine eindeutige Lösung ist. (Vgl. Guilford (1967b), S. 138).

⁴⁹Vgl. Guilford (1967b), S. 143.

⁵⁰Vgl. Torrance (1974b), S. 12.

⁵¹Vgl. Guilford (1967b), S. 138; Torrance (1974b), S. 56.

⁵²Vgl. Torrance (1974b), S. 56f.; Torrance (1974a), S. 31-36.

besteht nach Torrance (1974a) ebenfalls die Möglichkeit die drei Scores aufzuaddieren.⁵³ Jedoch wird ebenso kritisch angemerkt, dass ein einzelner Score irreführend sein kann, da eine Testperson möglicherweise eine große Anzahl von Ideen produzieren kann, die aber weder eine hohe Originalität noch Flexibilität aufweisen.⁵⁴

2.2. Feedback

Feedback ist im organisatorischen Kontext ein weit verbreitetes und allgemein anerkanntes Instrument zur Einflussnahme auf Motivation und Verhalten der MitarbeiterInnen.⁵⁵ In der Forschungsliteratur gibt es dennoch vielfältige Bedeutungsnuancen bei der Abgrenzung des Begriffs.⁵⁶ Die ausformulierten Definitionen auf die sich die vorliegende Arbeit bezieht, lassen sich in Tabelle 1 im Anhang finden.

Ein Minimalkonsens unter den verschiedenen AutorInnen ist die Feststellung, dass Feedback Informationen beinhaltet, die sich als Rückmeldung auf das Verhalten von EmpfängerIn verstehen lassen.⁵⁷ Ilgen et al. (1979) bezeichnen daher Feedback als eine spezielle Form des allgemeinen Kommunikationsprozesses (SenderIn übermittelt Nachricht an EmpfängerIn).⁵⁸ Wie die Informationen des Feedbacks entstehen, wird nur bei einigen AutorInnen aufgegriffen, jedoch immer ähnlich beschrieben. So vergleicht die Feedback sendende Person das Verhalten von dem/der betroffenen EmpfängerIn mit einem festgelegten Ziel oder Standard, um Informationen über die Abweichung zu generieren.⁵⁹ Differenzen ergeben sich aus der Frage, welche Art von Informationen wirksames Feedback enthalten muss. Hierbei wird häufig Bezug auf Kulhavy (1977) genommen, der von einem Kontinuum zwischen dem sogenannten knowledge of results (simple Auskunft über Richtig oder Falsch bei einer Aufgabe oder Verhalten⁶⁰) bis hin zu korrigierenden Informationen ausgeht.⁶¹ In einem weiteren Beitrag von Kulhavy und Stock (1989) wird analog die dichotome Beurteilung von Richtig oder Falsch als verification und die substantiellen, korrigierenden Informationen als elaboration bezeichnet.⁶² Kluger und DeNisi (1996) erläutern den Unterschied zwischen den beiden Komponenten anhand des folgenden Beispiels:

„A KR [knowledge of results - Anm. d. Verf.] intervention may include, for example, a message that 'your average typing speed is 100 words per

minute'. Yet, providing information about how one performs a job, such as 'you do not use your thumb for typing', is not a KR intervention because it does not contain information about the effectiveness of one's performance [...]. Yet, such an intervention is an FI [feedback intervention - Anm. d. Verf.] and therefore is included in our definition.“⁶³

Mit Bezug auf die Frage, welche Komponenten in wirksamen Feedback enthalten sein sollen, beziehen die meisten AutorInnen die Position, dass reines knowledge of results noch nicht ausreicht, sondern weiterführende Informationen vonnöten sind.⁶⁴ Nichtsdestotrotz gibt es ebenso AutorInnen, die Feedback nur mit knowledge of results gleichsetzen.⁶⁵

Unabhängig davon, welche Art von Informationen das Feedback übermittelt, ist die Art wie die Informationen mitgeteilt werden abhängig von Bestandteilen des Feedbacks, die unterschiedlich ausgeprägt sein können. Zwei für die vorliegende Arbeit maßgebliche Komponenten sind das Feedbackvorzeichen und der Feedbackstil. Das Feedbackvorzeichen ist das positive oder negative Ergebnis aus dem Vergleich der Performance eines Individuums mit situationsbedingten oder normativen Kriterien.⁶⁶ Feedback mit einem positiven Vorzeichen ist also ein Indikator für eine Leistung die das Kriterium übertrifft, entsprechendes gilt für negatives Feedback.⁶⁷ Der Feedbackstil ergibt sich aus der Formulierung des Feedbacks und wird meistens als kontrollierend oder informativ beschrieben.⁶⁸ Kontrollierendes Feedback zeichnet sich, unabhängig vom Vorzeichen, dadurch aus, dass äußere Zwänge bei der Formulierung deutlich sind. So wird Druck aufgebaut, dass ein bestimmtes Ziel erreicht werden muss und der Anspruch der Feedback-gebenden Person steht im Vordergrund.⁶⁹ Informatives Feedback hingegen stellt die Feedback-empfangende Person in den Fokus und vermittelt Informationen und Ratschläge ohne ein zu erreichendes Ziel zu formulieren. Dadurch soll bewirkt werden, dass der/die Feedback-EmpfängerIn sich selbst in der Kontrolle ihres/seines Verhaltens sieht.⁷⁰

Ebenso wie keine durchgängige Definition von Feedback existiert, besteht auch über die Auswirkung von Feedback auf die Leistung von Personen keine einheitliche Auffassung. Die kognitiven Mechanismen lassen sich anhand der sogenannten Feedback-Standard Gap erläutern. Diese entsteht durch die Differenz zwischen dem aktuellen Niveau (Information aus Feedback) und dem erwünschten Niveau der Leistung

⁵³Vgl. Torrance (1974b), S. 56.

⁵⁴Vgl. Torrance (1974b), S. 57.

⁵⁵Vgl. Ilgen et al. (1979), S. 349.

⁵⁶Vgl. Fengler (2010), S. 6.

⁵⁷Vgl. u. a. Ammons (1956), S. 279; Ashford und Cummings (1983), S. 372; Carson und Carson (1993), S. 42; Hattie und Timperley (2007), S. 81; Ilgen et al. (1979), S. 350; Kluger und DeNisi (1996), S. 255; Kulhavy (1977), S. 212; Kulhavy und Stock (1989), S. 285; Otley (1999), S. 369; Ramaprasad (1983), S. 4; Shute (2008), S. 154.

⁵⁸Vgl. Ilgen et al. (1979), S. 350.

⁵⁹Vgl. Ashford und Cummings (1983), S. 372; Hattie und Timperley (2007), S. 82; Otley (1999), S. 369; Ramaprasad (1983), S. 5.

⁶⁰Vgl. Reber und Reber (2001), S. 381.

⁶¹Vgl. Kulhavy (1977), S. 212.

⁶²Vgl. Kulhavy und Stock (1989), S. 285.

⁶³Kluger und DeNisi (1996), S. 255.

⁶⁴Vgl. Hattie und Timperley (2007), S. 82; Kluger und DeNisi (1996), S. 255; Kulhavy (1977), S. 212; Kulhavy und Stock (1989), S. 285; Shute (2008), S. 157f.

⁶⁵Vgl. Ammons (1956), S. 279; Carson und Carson (1993), S. 37.

⁶⁶Vgl. Kluger und DeNisi (1996), S. 259; Zhou (1998), S. 262.

⁶⁷Vgl. Zhou (2008), S. 130.

⁶⁸Vgl. Deci (1975), S. 142; Ryan (1982), S. 451; Zhou (1998), S. 263; Benzer (2006), S. 11.

⁶⁹Vgl. Ryan (1982), S. 451; Zhou (1998), S. 263.

⁷⁰Vgl. Zhou (2008), S. 130.

(Standard der erreicht werden soll).⁷¹ Im Sinne der Zielsetzungstheorie von Latham und Locke (1991), wird ein Individuum bei einer positiven Differenz (Leistung übersteigt Standard) das Verhalten beibehalten. Bei einer negativen Differenz (Leistung unterschreitet Standard), wird das Individuum, wenn es unzufrieden damit ist und Zuversicht hat sich zu verbessern, versuchen die Leistung in Zukunft zu erhöhen um dem Standard zu entsprechen.⁷² Allerdings merken Kluger und DeNisi (1996) in ihrer Feedback Intervention Theory an, dass nur bestimmte Feedback-Standard Gaps beachtet und reguliert werden, da die Aufmerksamkeit von Individuen begrenzt ist.⁷³

Bezüglich der Umsetzung der Informationen des Feedbacks, gibt es keine eindeutigen Erkenntnisse. Zwar beschrieb Ammons bereits 1956 die überwiegend positiven Effekte von Feedback auf Lernprozesse und Motivation mit Verweis auf diverse empirische Studien.⁷⁴ Laut Kluger und DeNisi (1996) ist Ammons Review und deren fortwährende Verbreitung⁷⁵ als Evidenz für die positiven Effekte von Feedback jedoch auf Grund der Vernachlässigung inkonsistenter empirischer Daten und methodischer Mängel zu kritisieren.⁷⁶ Vielmehr belegen die Autoren anhand einer Metaanalyse, dass die Effekte von Feedback-Interventionen sehr variabel sind. So hatte ein Drittel der untersuchten Feedback-Interventionen sogar einen negativen Effekt auf die Leistung von Versuchspersonen (VPn).⁷⁷ Als Gründe nennen Kluger und DeNisi, dass unterschiedliche Feedbackarten die Aufmerksamkeit von Personen verlagern kann und sich von einer förderlichen Aufmerksamkeit auf die Aufgabe hin zu einer nicht förderlichen Aufmerksamkeit auf das Selbst verschieben kann.⁷⁸ Auch Latham und Locke (1991) äußern sich kritisch gegenüber der undifferenzierten, positiven Einschätzung und Verwendung von Feedback:

„[...] feedback is only information, that is, data, and as such has no necessary consequences at all. [...] its effect on action depends on how it is appraised and what decisions are subsequently made with respect to it.”⁷⁹

Ilgen et al. (1979) verweist weiterhin darauf, dass die Wirkung von Feedback pauschal nicht beschrieben werden kann, sondern von unterschiedlichen Bedingungen determiniert wird. Wahrnehmung und Reaktion der empfangenden Person auf Feedback sind demnach abhängig von Persönlichkeitsmerkmalen der sendenden sowie empfangenden Person, wie auch von der Art der Nachricht selbst.⁸⁰

2.3. Selbstwirksamkeit

Selbstwirksamkeit wird von Bandura (1998) als „beliefs in one's capabilities to organize and execute the courses of actions required to produce given attainments“⁸¹ definiert und hat seine Wurzeln in der von ihm entwickelten sozial-kognitiven Lerntheorie. Gemäß dieser steuern subjektive Überzeugungen, wie einerseits Handlungs-Ergebnis-Erwartungen und andererseits Selbstwirksamkeitserwartungen kognitive, emotionale, motivationale und aktionale Prozesse.⁸² Eine Handlungs-Ergebnis-Erwartung bezeichnet die Einschätzung einer Person, dass ein gewisses Verhalten zu einem bestimmten Resultat führt. Eine Selbstwirksamkeitserwartung hingegen ist die Überzeugung einer Person, in der Lage zu sein, das benötigte Verhalten zur Erreichung des Resultats auszuführen (Siehe Abbildung 3). Demgemäß sind Selbstwirksamkeits- und Handlungs-Ergebnis-Erwartung in dem Sinne voneinander abgegrenzt, dass eine Person zwar wissen kann, welches Verhalten zu einem bestimmten Resultat führt, jedoch durch Zweifel an den eigenen Fähigkeiten nicht in der Lage dazu ist, dieses Verhalten auch selbst durchzuführen.⁸³

Eine ausgeprägte Selbstwirksamkeitserwartung beeinflusst sowohl grundsätzlich das Denken, Fühlen und Handeln als auch die Zielsetzung, Anstrengung und Ausdauer von Personen.⁸⁴ Das bedeutet, dass eine Person mit erhöhter Selbstwirksamkeitserwartung eine komplizierte Aufgabe eher als Herausforderung angeht die intrinsisches Interesse weckt, statt sie als Bedrohung wahrzunehmen.⁸⁵ Infolgedessen, setzt sie sich höhere Ziele und verfolgt diese in Erwartung auf Erfolg mit Ausdauer und Anstrengung.⁸⁶ Der Grad der Selbstwirksamkeit bestimmt auch, wie sehr sich Personen bei der Aufgabenerfüllung bemühen und wie lange sie diese Bemühungen im Angesicht von Hindernissen und negativen Erfahrungen aufrechterhalten.⁸⁷ Sollten doch Rückschläge und Misserfolge eine Person mit starker Selbstwirksamkeitserwartung von der Erfüllung der Zielsetzung abhalten, wird dies nicht auf die Person und ihre Fähigkeiten an sich zurückgeführt, sondern entweder auf externe Umstände oder variable Faktoren wie zu wenig Bemühen oder Wissen.⁸⁸

Im Folgenden werden die vier von Albert Bandura formulierten Einflussquellen vorgestellt, die den Glauben von Personen in ihre Selbstwirksamkeit beeinflussen, wobei für die vorliegende Arbeit vor allem die verbale Überzeugung von Bedeutung sein wird. Die erste und effektivste Art eine robuste Selbstwirksamkeitserwartung aufzubauen, ist das Erleben von Erfolgserlebnissen, die sich den Anstrengungen und Fähigkeiten der Person zuschreiben lassen.⁸⁹ Dabei sind

⁷¹Vgl. Shute (2008), S. 157.

⁷²Vgl. Latham und Locke (1991), S. 226.

⁷³Vgl. Kluger und DeNisi (1996), S. 259.

⁷⁴Vgl. Ammons (1956), S. 283-286.

⁷⁵Vgl. CASAS-ARCE et al. (2017), S. 7; Fombelle et al. (2016), S. 688; Froehlich et al. (2014), S. 511; Qian et al. (2016), S. 479.

⁷⁶Vgl. Kluger und DeNisi (1996), S. 256f..

⁷⁷Vgl. Kluger und DeNisi (1996), S. 275.

⁷⁸Vgl. Kluger und DeNisi (1996), S. 275.

⁷⁹Latham und Locke (1991), S. 224.

⁸⁰Vgl. Ilgen et al. (1979), S. 350.

⁸¹Bandura (1998), S. 421.

⁸²Vgl. Schwarzer und Jerusalem (2002), S. 35.

⁸³Vgl. Bandura (1977), S. 193.

⁸⁴Vgl. Schwarzer und Jerusalem (2002), S. 37.

⁸⁵Vgl. Bandura (1977), S. 194.

⁸⁶Vgl. Schwarzer und Jerusalem (2002), S. 37.

⁸⁷Vgl. Bandura (1977), S. 194.

⁸⁸Vgl. Bandura (1998), S. 421.

⁸⁹Vgl. Schwarzer und Jerusalem (2002), S. 42.



Abbildung 1: Schematische Darstellung des Unterschieds zwischen Selbstwirksamkeitserwartung und Handlungs-Ergebnis-Erwartung. Quelle: Eigene Darstellung nach Bandura (1997), S. 22.

nicht nur die mühelosen Erfolge maßgeblich, sondern vielmehr die Erfahrung, Hindernisse durch ausdauernde Bemühungen überwinden zu können.⁹⁰ Eine weitere Möglichkeit, Selbstwirksamkeit zu stärken, ermöglicht die stellvertretende Erfahrung, welche auch Beobachtungs- oder Modelllernen genannt wird.⁹¹ Diese vollzieht sich durch die Beobachtung von ähnlichen Personen bei deren erfolgreicher Erreichung bestimmter Ziele. Die beobachtende Person fühlt sich dadurch in ihrer eigenen Fähigkeit gestärkt ähnliche Ziele zu erreichen.⁹² Eine dritte, zusätzliche Möglichkeit den Glauben an die eigenen Fähigkeiten zu stärken, ist die verbale Überzeugung durch Andere. Personen können verbal davon überzeugt werden, dass sie die Fähigkeiten haben, bestimmte Aufgaben zu meistern. Infolgedessen mobilisieren sie häufig mehr Ressourcen und schaffen es Hürden bei der Aufgabenlösung zu überkommen und dabei Erfolge zu erzielen, was in einer erhöhten Selbstwirksamkeit resultiert.⁹³ Der vierte und schwächste Impuls zur Beeinflussung der Selbstwirksamkeitserwartung besteht in der emotionalen Erregung. Die Beurteilung der eigenen Fähigkeiten kann von der aktuellen Gemütslage mitbestimmt werden. So ist es unerlässlich, dass eine Person lernt ihre emotionalen und physischen Reaktionen richtig wahrzunehmen und zu interpretieren, damit diese entweder nicht als Hinweis auf mangelnde Kompetenz gedeutet werden oder Fertigkeiten erworben werden, um körperliche Reaktionen in schwierigen Situationen unter Kontrolle zu bringen.⁹⁴

2.4. Theoretische Zusammenhänge von Feedback, Selbstwirksamkeit und Kreativität

Dieser Abschnitt dient dazu die Kausalzusammenhänge zwischen Feedback, Selbstwirksamkeit und Kreativität zu betrachten. Dafür wird zuerst die Beziehung zwischen Selbstwirksamkeit und Kreativität sowie Feedback und Kreativität getrennt voneinander untersucht. Daraufaufgehend soll die indirekte Wirkung von Feedback über Selbstwirksamkeit auf Kreativität literaturbasiert erarbeitet werden. Hiermit werden die Zusammenhänge dargestellt, die in Kapitel 3 anhand

dem experimentellen Stand der Forschung auf ihre Gültigkeit überprüft werden.

Bei dem Konstrukt Selbstwirksamkeit wird Bezug auf Kreativität in verschiedenen Formen genommen. Grundsätzlich bezieht sich der Glaube an die eigenen Fähigkeiten nicht auf Aufgaben, die routinemäßig lösbar sind, sondern vielmehr auf solche, die eine schwierige Anforderungssituation für eine Person darstellen, ganz ähnlich den kreativen Erforderlichkeiten.⁹⁵ Im Übrigen beschreibt schon Bandura (1997), dass eine starke Selbstwirksamkeitserwartung eine Bedingung für kreatives Arbeiten darstellt, weil diese ein starkes Ausmaß an Beharrlichkeit fördert und Bewältigungsanstrengungen steigert, was für kreative Aufgaben und den damit verbundenen Investitionen in Zeit und Anstrengungen fundamental ist.⁹⁶ Weiterhin gehen erhöhte Selbstwirksamkeitsüberzeugungen mit bestimmten kognitiven Komponenten einher wie beispielsweise eine breiter aufgestellte Informationssuche und eine bessere Möglichkeit sein Gedächtnis abzurufen, was ebenfalls notwendig für den kreativen Prozess ist.⁹⁷ Darüber hinaus haben Tierney und Farmer (2002) basierend auf dem Konstrukt der allgemeinen Selbstwirksamkeit, kreative Selbstwirksamkeit (creative self-efficacy - CSE) als eigenständiges Selbstkonzept eingeführt und als „the belief one has the ability to produce creative outcomes“⁹⁸ definiert. Den Mehrwert von CSE angesichts des schon begründeten Zusammenhangs zwischen der allgemeinen Selbstwirksamkeit und Kreativität, legitimieren Tierney und Farmer durch Elemente, die im kreativen Prozess unabdingbar sind, jedoch im Rahmen der allgemeinen Selbstwirksamkeit nicht auftreten.⁹⁹ Diese Elemente sind das was Amabile in ihrem Komponentenmodell der Kreativität als kreativitätsspezifische Prozesse bezeichnet. Diese sind unter anderem, die Fähigkeit mentale Sets zu sprengen, die Kenntnis von Heuristiken zur Generierung von neuartigen Ideen, den Mut unkonventionelle Perspektiven einzunehmen und das Loslösen vom Streben nach gesellschaftlicher Akzeptanz.¹⁰⁰ Tierney und Farmer argumentieren, dass Personen, die sich ihrer kreativen Fähigkeiten bewusst sind, aller

⁹⁰Vgl. Bandura (1998), S. 422.

⁹¹Vgl. Barysch (2016), S. 204.

⁹²Vgl. Luthans et al. (2007), S. 37.

⁹³Vgl. Bandura (1998), S. 422.

⁹⁴Vgl. Bandura (1998), S. 422.

⁹⁵Vgl. Schwarzer und Jerusalem (2002), S. 35.

⁹⁶Vgl. Bandura (1997), S. 239.

⁹⁷Vgl. Cervone et al. (1991), S. 264.

⁹⁸Tierney und Farmer (2002), S. 1138.

⁹⁹Vgl. Tierney und Farmer (2002), S. 1140.

¹⁰⁰Vgl. Amabile (1996), S. 88-90.

Voraussicht nach auch die von Amabile benannten Kompetenzen besitzen. Der Grund dies anzunehmen, liegt in der vorrangigen Einflussquelle der Selbstwirksamkeit: dem Erleben von Erfolgserlebnissen. Daher müssen gewisse kreative Kenntnisse, vormals zu Erfolg geführt haben und beweisen im Zirkelschluss die Existenz von oben genannten kreativitätsspezifischen Prozessen.¹⁰¹

Die Beziehung von Feedback und Kreativität wird von Zhou (2008) als „unlikely event“¹⁰² bezeichnet, da einerseits Feedback im Arbeitsumfeld weit verbreitet und häufig genutzt wird und andererseits Kreativität als erstrebenswert aber schwer fassbar angesehen wird. Dass durch ein so gebräuchliches Instrument wie Feedback, ein so ungewöhnliches Phänomen wie Kreativität gefördert werden kann, scheint deshalb zuerst ungewöhnlich.¹⁰³ Dennoch gibt es in der Forschungsliteratur zwei wesentliche Gründe, die die Wirkung von Feedback auf Kreativität begründen: die intrinsische Motivation sowie die Informationen über die Anforderungen der Aufgaben. Feedback steigert unter bestimmten Bedingungen die intrinsische Aufgabenmotivation und damit die Kreativität von Personen.¹⁰⁴ Den Prozess, wie Feedback sich auf die intrinsische Motivation auswirkt, beschreiben Zhou und George (2001) mit Bezug auf Utman (1997) und Amabile (1983b) folgendermaßen. Feedback, welches sich auf eine Aufgabe bezieht, lenkt die Aufmerksamkeit der Person auf die Aufgabe und fördert damit das Interesse an der Aufgabe selbst und hilft somit die Motivation, die Leistung durch Lernen und Weiterentwicklung zu steigern.¹⁰⁵ Eine Bedingung für die positive Wirkung ist dabei der Stil, in dem Feedback übermittelt wird. So schlussfolgert Deci (1975), dass kontrollierendes Feedback das Interesse an der Aufgabe mindert, indem sich das Individuum nicht mehr auf die Aufgabe konzentriert, sondern sich über das externe Feedback und sich selbst Gedanken macht.¹⁰⁶ Informatives Feedback liefert jedoch synergetische Effekte auf die intrinsische Motivation.¹⁰⁷ Auch in weiteren Abhandlungen zu diesem Thema wird Bezug auf die intrinsische Motivation und diese Argumentationskette genommen.¹⁰⁸ Darüber hinaus, agiert Feedback als Quelle von Informationen über die Ziele und Anforderungen der kreativen Arbeit.¹⁰⁹ Dies ist insbesondere im kreativen Kontext wichtig, weil Informationen nicht aus der Aufgabenumgebung gezogen werden können, wie es bei Quantitätsaufgaben der Fall ist. Dort können Personen leichter selbst einschätzen, auf welchem Level sie sich befinden. Da Kreativität jedoch in keiner linearen Beziehung zum Bemühen steht und schwer selbst zu evaluieren ist, ist Feedback

besonders hilfreich, um die eigene Leistung einschätzen und verbessern zu können.¹¹⁰

Einen indirekten Effekt übt Feedback über die verbale Überzeugung auf Selbstwirksamkeit bzw. die wahrgenommene Kompetenz¹¹¹ aus, welche sich wiederum positiv auf Kreativität auswirkt. Die verbale Überzeugung ist eine von vier Einflussquellen, die Personen in ihrer Selbstwirksamkeit bekräftigen oder schwächen kann. Feedback kann diese überzeugenden Informationen enthalten.¹¹² Die Bedingungen unter denen eine positive Wirkung auf die Selbstwirksamkeit beobachtbar ist, hängen vom Feedbackvorzeichen ab. Die relevante Forschungsliteratur ist sich einig, dass positives Feedback per se die Selbstwirksamkeit steigert, da es das Individuum in seinen Fähigkeiten bestärkt.¹¹³ In Bezug auf negatives Feedback, wird angenommen, dass dieses grundsätzlich mit einer Verminderung von Selbstwirksamkeit einhergeht, da es mangelnde Leistungsfähigkeit signalisiert.¹¹⁴

Die verschiedenen Wirkungszusammenhänge sind in Abbildung 3 zusammengefasst und lassen sich folgendermaßen komprimieren. Feedback hat durch die Bereitstellung von Informationen über die Anforderungen von kreativen Aufgaben eine direkte, positive Wirkung auf Kreativität. Darüber hinaus hat Feedback, unter der Bedingung, dass es in einem informativen Stil übermittelt wird, eine positive Wirkung auf Kreativität durch eine Steigerung der intrinsischen Motivation. Die indirekte Wirkung von Feedback vollzieht sich über die Selbstwirksamkeit einer Person. Unter der Bedingung, dass Feedback ein positives Vorzeichen hat, operiert dieses als verbale Überzeugung und steigert damit die Selbstwirksamkeit, welche wiederum durch erhöhte kognitive Leistung und Kenntnis von kreativitätsspezifischen Prozessen einen positiven Einfluss auf die Kreativität hat.

3. Zum Erkenntnisstand der experimentellen Forschung über den Zusammenhang von Feedback, Selbstwirksamkeit und Kreativität

Das folgende Kapitel widmet sich der empirischen Überprüfung der im vorherigen Kapitel dargestellten theoretischen Kausalzusammenhänge durch ausgewählte experimentelle Studien. Da dem indirekten Effekt von Feedback über Selbstwirksamkeit auf Kreativität bisher keine Aufmerksamkeit in der experimentellen Forschung gewidmet wurde, werden im Folgenden nur Experimente über Feedback und Kreativität sowie Selbstwirksamkeit und Kreativität präsentiert. Schließlich werden im Hinblick auf die gefundenen

¹⁰¹Vgl. Tierney und Farmer (2002), S. 1140.

¹⁰²Zhou (2008), S. 125.

¹⁰³Vgl. Zhou (2008), S. 125.

¹⁰⁴Vgl. Zhou (2008), S. 128-130.

¹⁰⁵Vgl. Zhou und George (2001), S. 685; Utman (1997), S. 172; Amabile (1983a), S. 369.

¹⁰⁶Vgl. Deci (1975), S. 142.

¹⁰⁷Vgl. Amabile (1996), S. 118.

¹⁰⁸Vgl. u. a. Amabile und Gryskiewicz (1987), S. 22; Benzer (2006), S. 14; Zhou (1998), S. 262.

¹⁰⁹Vgl. Byron und Khazanchi (2012), S. 811f.; Zhou (2008), S. 129.

¹¹⁰Vgl. Carson und Carson (1993), S. 37.

¹¹¹Die wahrgenommene Kompetenz wird als "individual's perception that he or she is capable of doing a task" (Zhou (1998), S. 263) definiert und ist Bestandteil der Cognitive Evaluation Theory von Deci und Ryan (1985). Im Folgenden wird das Konzept der wahrgenommenen Kompetenz, einvernehmlich mit der bestehenden Forschungsliteratur, synonym mit dem Konzept der Selbstwirksamkeit verwendet (Vgl. Rodgers et al. (2014), S. 6f.).

¹¹²Vgl. Bandura (1997), S. 101.

¹¹³Vgl. Bandura (1997), S. 102; Benzer (2006), S. 13f.; Dimotakis et al. (2017), S. 1516; Latham und Locke (1991), S. 226; Zhou (1998), S. 264.

¹¹⁴Vgl. Benzer (2006), S. 13f.; Byron und Khazanchi (2012), S. 812; Dimotakis et al. (2017), S. 1516; Zhou (1998), S. 263f.

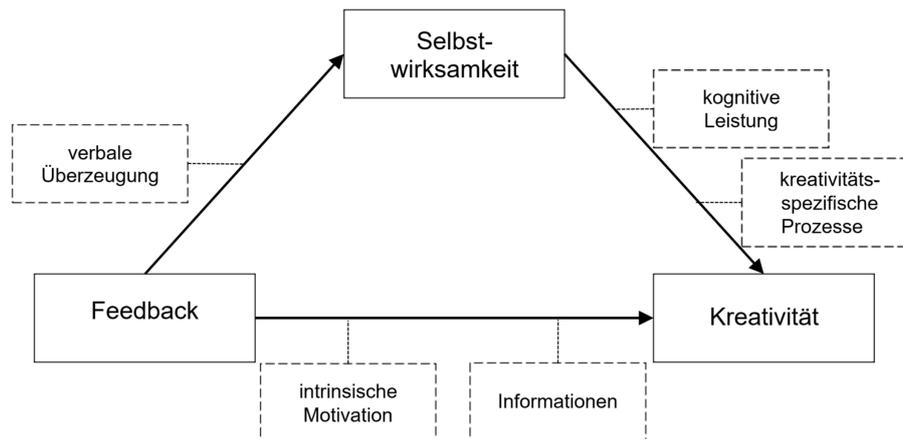


Abbildung 2: Wirkungszusammenhänge von Feedback, Selbstwirksamkeit und Kreativität. Quelle: Eigene Darstellung.

Limitationen, Lücken für die folgenden eigenständig konzipierten experimentellen Designs skizziert.

3.1. Feedback und Kreativität

Obwohl die Nutzung von Feedback weit verbreitet und Kreativität in fast allen organisatorischen Bereichen von großer Bedeutung ist, wurde bisher wenig Forschung zu der Beziehung von Feedback und Kreativität betrieben.¹¹⁵ Für die vorliegende Arbeit werden vier relevante Experimente von Benzer (2006), Fodor und Carver (2000), Ryan (1982) und Zhou (1998)¹¹⁶ erläutert und kritisch bewertet. Dies geschieht indem Gemeinsamkeiten und Unterschiede bei der Wahl der Aufgabe und der Messgrößen, des Verfahrens und der Treatments beschrieben und Limitationen herausgearbeitet werden. Die Experimente von Ryan (1982) und Zhou (1998) wurden dabei insbesondere ausgewählt, da diese seit ihrem Erscheinen eine hohe Reichweite haben. Fodor und Carver (2000) sowie Benzer (2006) sind im Vergleich weniger einflussreich, indes ebenso interessant. Fodor und Carver (2000) beziehen sich nämlich, anders als die Meisten, auf eine Aufgabe, bei der bereichsrelevante Fähigkeiten ebenso eine Rolle spielen wie kreativitätsrelevante Prozesse. Benzer (2006) hingegen orientiert sich an dem Vorgehen von Zhou (1998), kommt jedoch durch Limitationen in seiner Studie zu anderen Ergebnissen.

Mit Ausnahme von Ryan (1982) nutzen alle AutorInnen heuristische, offen gestellte Aufgaben, um die Kreativität der VPn zu beobachten. Zhou (1998) und Benzer (2006) wenden eine Variation der gleichen Aufgabe von Shalley (1991) an.¹¹⁷ Die Aufgabe ist eine sogenannte Postkorb-Fallstudie¹¹⁸, bei der sich die VPn in die Rolle eines/r Perso-

nalcheffn eines Stahl verarbeitenden Betriebs hineinversetzen und kreative, d.h. neuartige und nützliche Lösungen zu einer Reihe von Problemen finden müssen.¹¹⁹ Die Probleme sind in 22 Memos erläutert und betreffen Themen wie Vorfälle der sexuellen Belästigung am Arbeitsplatz, Diebstähle durch MitarbeiterInnen und Produktivitätsschwäche.¹²⁰ Fodor und Carver (2000) stellen zwei technische Probleme vor, die möglichst originell zu lösen und für die VPn, die hauptsächlich Ingenieurwissenschaften studieren, angemessen sind.¹²¹ Diese Aufgaben sind deshalb heuristisch- komplex, weil der Lösungsweg nicht bekannt ist, sondern von dem/der AufgabenlöserIn entwickelt werden muss. Die Aufgaben sind nützlich, um Feedback zu geben, da die VP bei offen gestellten Fragen schwerer abschätzen kann, wie gut seine/ihre kreative Leistung ausgefallen ist und angewiesen auf eine Rückmeldung ist.¹²² Die kreative Leistung wird bei den oben benannten Aufgaben durch die CAT von unabhängigen RichterInnen bewertet.¹²³ Ryan (1982) hingegen verwendet eine algorithmische Aufgabe mit einem bekannten Lösungsweg, die sogenannte hidden figures Aufgabe. Hier soll die VP versuchen, möglichst häufig den Namen „Nina“ in einer Comic-Zeichnung zu finden.¹²⁴ Die Wahl der algorithmischen Aufgabe lässt sich dadurch erklären, dass seine Studie nicht Kreativität, sondern intrinsisches Interesse der VPn als abhängige Variable untersucht. Für die vorliegende Arbeit sind Ryans Studien trotzdem relevant, da das intrinsische Interesse als Vorläufer von Kreativität behandelt wird und sich Autorinnen wie Zhou (1998) explizit auf seine Forschung beziehen.

Der Ablauf der vier Experimente zeigt viele Gemeinsamkeiten und wird in Abbildung ?? beispielhaft dargestellt. Es lässt sich beobachten, dass alle experimentellen Studien

¹¹⁵Vgl. Beckers et al. (2010), S. 476.

¹¹⁶Überblick in Tabelle 4, Anhang.

¹¹⁷Vgl. Benzer (2006), S. 17; Zhou (1998), S. 267.

¹¹⁸Eine Postkorb-Fallstudie ist ein, in Assessment-Centern beliebtes Testverfahren, um die Effektivität und Effizienz einer BewerberIn unter Zeitdruck zu testen. Dazu wird der Person in einer Simulation ein beispielhaftes Posteingangsfach einer ManagerIn präsentiert, das unterschiedliche, alltägliche Probleme enthält, die Aktion oder Delegation erfordern (Vgl. Gill (1979), S. 185).

¹¹⁹Vgl. Shalley (1991), S. 181.

¹²⁰Vgl. Zhou und Shalley (2003), S. 176.

¹²¹Vgl. Fodor und Carver (2000), S. 384, 386f..

¹²²Vgl. Carson und Carson (1993), S. 37.

¹²³Vgl. Benzer (2006), S. 19; Fodor und Carver (2000), S. 383; Zhou (1998), S. 268.

¹²⁴Vgl. Ryan (1982), S. 454.



Abbildung 3: Messwiederholungen und Ablaufplan der ausgewählten Experimente. Quelle: Eigene Darstellung.

ein between-subjects design verwenden, also verschiedene Personen unterschiedlichen Treatments zugeordnet werden.¹²⁵ Bei allen Studien gibt es eine Kontrollgruppe, die kein Feedback-Treatment bekommt. Ebenso nutzen alle AutorInnen ein zweistufiges Verfahren, bei dem zuerst eine der oben beschriebenen Kreativitätsaufgaben gelöst werden soll. Darauf folgend erhalten die VPn je nach Treatment Feedback zu ihrer Leistung und sollen danach eine zweite Kreativitätsaufgabe lösen. Bei beiden Kreativitätsaufgaben wird die abhängige Variable (AV) Kreativität gemessen. Die zwei Aufgaben sind sich bei allen AutorInnen ähnlich, damit die Informationen des Feedbacks sich ebenso auf die nachfolgende Aufgabe übertragen lassen können. Dies wird jedoch durch den mangelnden Informationsgehalt des Feedbacks zunichtegemacht. Das Feedback beschränkt sich in den ausgewählten Experimenten ausschließlich auf die verification Komponente. Dies wird bei Betrachtung der Formulierung des Feedbacks deutlich. „You did really well. Congratulations! Keep up the good work.“¹²⁶ bzw. „You didn’t do very well, sorry.“¹²⁷ wurde bei Zhou (1998) als informatives positives bzw. negatives Feedback kommuniziert.¹²⁸ Das exemplarische Feedback zeigt, dass den VPn, unabhängig vom Treatment, nur mitgeteilt wird, ob sie gut oder schlecht abgeschnitten haben, nicht aber, warum sie ein gutes oder schlechtes Ergebnis erzielt haben. Daraus ergibt sich die Schwierigkeit, dass die VPn durch das Feedback keine Informationen über die Anforderungen der Aufgabe erhalten und damit der direkte Wirkungszusammenhang zwischen Feedback und Kreativität möglicherweise nicht ausgeführt wird. Außerdem schadet die Eindimensionalität der Glaubwürdigkeit des Feedbacks, was in diesem Fall besonders problematisch ist, da das Feedback eine Täuschung ist, weil die VPn unabhängig von ihrer Leistung und nur gemäß der zufällig eingeteilten Versuchsbedingung Feedback bekommen. Um Glaubwürdigkeit zu simulieren, wird vor den VPn behauptet, dass die kreative Leistung entweder von RichterInnen¹²⁹, gemäß einem Handbuch mit Richtlinien¹³⁰ oder durch ein Inhaltsanalyseprogramm eines Computers¹³¹ bewertet wurde. Ebendies führte bei Benzer (2006) jedoch zu Zweifeln an der Richtigkeit des Feedbacks.¹³²

Alle AutorInnen variieren mehrere unabhängige Variablen. Hier sollen nur zwei für die vorliegende Arbeit relevante Treatments vorgestellt werden: das Feedbackvorzeichen und der Feedbackstil. Dabei wird ersteres von Benzer (2006), Fodor und Carver (2000), sowie Zhou (1998) variiert, und zweiteres von Benzer (2006), Ryan (1982) und Zhou (1998). Das Feedbackvorzeichen wird entweder anhand eines guten oder schlechten Punktestands im Vergleich zu Ergebnissen aus einem Pretest¹³³ oder mithilfe eines positiv oder negativ formulierten Vermerks¹³⁴ manipuliert. Bei der Variation des Feedbackstils gibt es ebenfalls Überschneidungen. Ryan (1982) formuliert kontrollierendes Feedback mithilfe der Phrase you should, um die VP spüren zu lassen, dass er/sie nicht aus eigenem Antrieb zu handeln, sondern fremdbestimmt ist und unter dem Druck steht, ein bestimmtes Ergebnis zu erreichen.¹³⁵ Zhou (1998) und Benzer (2006) orientieren sich an diesem Vorgehen.¹³⁶ Informatives Feedback wird unterschiedlich geäußert. Während bei Ryan (1982) nur ein Vergleich vom Punktestand der VP mit Vergleichswerten gezeigt¹³⁷ wird, sind bei Benzer (2006) und Zhou (1998) der Punktestand um Kommentare ergänzt.¹³⁸

Die Ergebnisse der ausgewählten Studien lassen sich mit Bezug auf die verwendeten Treatments zusammenfassen. Mit Blick auf die Wirkung von unterschiedlichen Feedbackvorzeichen auf Kreativität, kommen Fodor und Carver (2000) zu dem Ergebnis, dass positives Feedback die nachfolgende kreative Leistung fördert. Auch negatives Feedback begünstigt Kreativität bei einigen VP, jedoch nur unter der Bedingung, dass diese leistungs-¹³⁹ und nicht machtmotiviert¹⁴⁰ sind.¹⁴¹ Zhou (1998) weist nach, dass VPn die positives Feedback erhielten im Durchschnitt die höchste kreative Leistung zeigten. Darauf folgten die VPn, die kein Feedback bekamen und schließlich die VPn im negativen Feedback-Treatment.¹⁴² Laut Kapitel 2.4 sollte jedoch selbst negatives Feedback aufgrund der vermittelten Infor-

¹³³Vgl. Benzer (2006), S. 18; Zhou (1998), S. 267.

¹³⁴Vgl. Fodor und Carver (2000), S. 386f.

¹³⁵Vgl. Ryan (1982), S. 451f.

¹³⁶Vgl. Benzer (2006), S. 19; Zhou (1998), S. 267.

¹³⁷Vgl. Ryan (1982), S. 454.

¹³⁸Vgl. Benzer (2006), S. 18; Zhou (1998), S. 267.

¹³⁹Die Leistungsmotivation beschreibt die „allgemeine und relativ überdauernde Tendenz, als wesentlich bewertete Aufgaben mit Energie und Ausdauer bis zum erfolgreichen Abschluss zu bearbeiten“ (Drever und Fröhlich (1972), S. 170).

¹⁴⁰„Machtmotiviertes Verhalten ist dadurch gekennzeichnet, dass Individuen andere dominieren, über sie bestimmen bzw. Einfluss (Einfluss, sozialer) auf sie ausüben.“ (Vgl. Wirtz (2014), S. 984).

¹⁴¹Vgl. Fodor und Carver (2000), S. 389.

¹⁴²Vgl. Zhou (1998), S. 271.

¹²⁵Vgl. Benzer (2006), S. 17; Fodor und Carver (2000), S. 386; Ryan (1982), S. 454; Zhou (1998), S. 266.

¹²⁶Zhou (1998), S. 267.

¹²⁷Zhou (1998), S. 267.

¹²⁸Die weiteren Feedbackphrasen befinden sich in Tabelle 5 und 6, Anhang.

¹²⁹Vgl. Benzer (2006), S. 17.

¹³⁰Vgl. Fodor und Carver (2000), S. 386.

¹³¹Vgl. Zhou (1998), S. 267.

¹³²Vgl. Benzer (2006), S. 63.

mationen über Anforderungen der Aufgabe zuträglicher für kreative Leistung sein, als überhaupt kein Feedback. Dieser Widerspruch lässt sich jedoch dadurch erklären, dass die Feedbackphrasen bei Zhou (1998), wie oben gezeigt, nicht besonders inhaltsreich waren und demnach kaum Informationen überbracht haben. Benzer (2006) kommt zu konträren Resultaten. Er weist in seiner Studie nach, dass in einem komplexen Aufgabenumfeld negatives Feedback zu einer höheren kreativen Leistung führt als positives Feedback.¹⁴³ Jedoch sind seine Erkenntnisse durch Limitationen seiner Studie beschränkt. So haben VPn die Glaubwürdigkeit seines Feedbacks angezweifelt und seine Feedbackformulierungen sind sehr abstrakt, so dass sie in unterschiedliche Richtungen interpretiert werden können.¹⁴⁴ Nichtsdestotrotz konnte die Hypothese, dass das Feedbackvorzeichen einen positiven Effekt auf die wahrgenommene Kompetenz hat, bestätigt werden.¹⁴⁵ Da die wahrgenommene Kompetenz bedeutungsähnlich mit Selbstwirksamkeit verwendet werden kann, bestätigt Benzers (2006) Feststellung, trotz Limitationen, den theoretisch aufgestellten Wirkungszusammenhang zwischen Feedback über Selbstwirksamkeit auf Kreativität. Im Hinblick auf die Wirkung des Feedbackstils kommen Ryan (1982) und Zhou (1998) auf ähnliche Ergebnisse und bestätigen die in Kapitel 2.4 dargestellten Zusammenhänge. Ein informativer Feedbackstil steigert intrinsisches Interesse der VPn¹⁴⁶ und führt bei VPn, in Verbindung mit einem positiven Feedbackvorzeichen, zu hohen Kreativitätswerten.¹⁴⁷

3.2. Selbstwirksamkeit und Kreativität

Den theoretischen Beweis, dass Selbstwirksamkeit eine positive Wirkung auf Kreativität hat, überprüfen die empirischen Studien von Beghetto et al. (2011), Brockhus et al. (2014), Lemons (2010), Pretz und McCollum (2014) und Tierney und Farmer (2002), Tierney und Farmer (2011).¹⁴⁸ Die Mehrheit der Untersuchungen verifiziert die Beziehung zwischen CSE und kreativer Performance dadurch, dass eine signifikante Korrelation festgestellt wird¹⁴⁹, Varianzen in Bewertungen von kreativer Performance durch CSE erklärt werden¹⁵⁰ und Kreativität als Funktion von CSE dargestellt wird¹⁵¹. Nichtsdestotrotz gibt es ebenso Studien, die infrage stellen, ob CSE und Kreativität zusammenhängen. So belegen zum Beispiel Pretz und McCollum (2014) anhand eines Experiments, dass CSE nicht mit der aktuellen kreativen Performance, sondern mit kreativen Erfolgen in der Vergangenheit korrelieren und, dass die Einschätzung der kreativen Leistung bei der konkreten Aufgabe in Bezug auf die tatsächliche kreative Performance weitaus stichhaltiger ist als die globale kreative Selbstwirksamkeit einer Person.¹⁵²

Die Ausprägung der kreativen Selbstwirksamkeit wird in den oben beschriebenen und weiterführenden Studien mit verschiedenen Fokussen auf CSE anhand von Skalen mit verschiedenen Items operationalisiert.¹⁵³ Der Großteil verwendet dazu entweder das 3-Item-Modell von Tierney und Farmer (2002)¹⁵⁴, die 3 Items von Beghetto (2006)¹⁵⁵ oder die 6 Items der Short Scale for Creative Self¹⁵⁶ von Karwowski (2012)¹⁵⁷ um die kreative Selbstwirksamkeit messen. Hass et al. (2016) nutzen sogar eine Kombination der erstgenannten Modelle von Beghetto und Tierney und Farmer.¹⁵⁸ Je nach Studie variiert der Wert von Cronbachs Alpha, er verweilt jedoch immer zwischen $\alpha = .62$ und $\alpha = .873$, womit sich die interne Konsistenz¹⁵⁹ der drei Skalen auf einem akzeptablen bis guten Level befindet.¹⁶⁰ Die Kreativität der VPn wird im Großteil der Studien mithilfe von Fragebögen, ausgefüllt von Vorgesetzten, LehrerInnen oder den VPn selbst, beurteilt. Diese umfassen bei Tierney und Farmer (2002), Tierney und Farmer (2011) entweder vier (2011) oder sechs (2002) Items¹⁶¹ über die kreativen Fähigkeiten der Angestellten und sind von Tierney et al. (1999) übernommen.¹⁶² Jedoch wird keine Angabe darüber gemacht, welche spezifischen Items übernommen werden. Bewertet wurde auf einer sechsstufigen Skala von 1 (= never) bis 6 (= always). Beghetto et al. (2011), bei denen LehrerInnen zur Bewertung der Kreativität ihrer SchülerInnen gebeten wurden, machen keine weiteren Angaben über ihre Fragebögen, außer dass sich die Bewertung auf die Fähigkeit bezieht, neuartige und angemessene Ideen zu artikulieren.¹⁶³ Lemons (2010) operationalisiert Kreativität anhand eines offenen Fragebogens über die selbstberichteten kreativen Fähigkeiten. Zuerst sollen die Studierenden ihre Kreativität auf einer sechsstufigen Likert-Skala bewerten und bei einem zweiten Item erklären, weshalb sie sich diese Bewertung gegeben haben.¹⁶⁴ Vorkehrungen gegenüber der Kritik, dass selbstberichtete Kriterien der Kreativität nicht akkurat sind¹⁶⁵, werden dadurch getroffen, dass die Studierenden ebenfalls Beispiele für ihre kreativen Tätigkeiten geben müssen, die objektiv überprüfbar sind.¹⁶⁶ In den beiden weiteren angeführten Studien wird Kreativität

¹⁵³Siehe Tabelle 2, Anhang.

¹⁵⁴Vgl. Jaussi et al. (2007), S. 252; Mathisen und Bronnick (2009), S. 25; Reiter-Palmon et al. (2012), S. 109f.; Richter et al. (2012), S. 1285; Tierney und Farmer (2002), S. 1141f.; Tierney und Farmer (2011), S. 282.

¹⁵⁵Vgl. Beghetto (2006), S. 450; Beghetto et al. (2011), S. 344; Pretz und McCollum (2014), S. 230.

¹⁵⁶Vgl. Karwowski et al. (2012).

¹⁵⁷Vgl. Karwowski (2012), S. 550; Puente-Díaz und Cavazos-Arroyo (2017), S. 5.

¹⁵⁸Vgl. Hass et al. (2016), S.439.

¹⁵⁹Die interne Konsistenz bezeichnet die interrelatedness einer Skala, somit das Ausmaß in dem die Fragen einer Skala miteinander korrelieren (Vgl. Cortina (1993), S. 100).

¹⁶⁰Vgl. George und Mallery (2003), S. 231.

¹⁶¹Siehe Tabelle 3, Anhang.

¹⁶²Vgl. Tierney et al. (1999), S. 620; Tierney und Farmer (2002), S. 1142; Tierney und Farmer (2011), S. 283.

¹⁶³Vgl. Beghetto et al. (2011), S. 344.

¹⁶⁴Vgl. Lemons (2010), S. 153.

¹⁶⁵Vgl. Reiter-Palmon et al. (2012), S. 108.

¹⁶⁶Vgl. Lemons (2010), S. 154.

¹⁴³Vgl. Benzer (2006), S. 28.

¹⁴⁴Vgl. Benzer (2006), S. 63f.

¹⁴⁵Vgl. Benzer (2006), S. 50.

¹⁴⁶Vgl. Ryan (1982), S. 456.

¹⁴⁷Vgl. Zhou (1998), S. 269.

¹⁴⁸Überblick in Tabelle 7, Anhang.

¹⁴⁹Vgl. Brockhus et al. (2014), S. 440.

¹⁵⁰Vgl. Beghetto et al. (2011), S. 347; Tierney und Farmer (2002), S. 1142.

¹⁵¹Vgl. Lemons (2010), S. 151; Tierney und Farmer (2011), S. 287.

¹⁵²Vgl. Pretz und McCollum (2014), S. 233.

mittels Brainstorming-Aufgaben, die anhand der Divergent Thinking Kategorien fluency, flexibility und originality¹⁶⁷ sowie fluency und originality¹⁶⁸ bewertet wurden, operationalisiert. Bei Pretz und McCollum (2014) werden darüber hinaus zwei, von den VPn ausgewählte, Antworten mithilfe der CAT von sechs fachkundigen RichterInnen auf einer sechsstufigen Skala bewertet.¹⁶⁹

Die Lücke für zwei eigene experimentelle Design ergibt sich zuerst aus der Gegebenheit, dass in keiner der ausgewählten Studien der Informationsgehalt des Feedbacks auf den elaboration Part ausgeweitet war. Förderliche Effekte durch einen Informationsgewinn über die Anforderungen der Aufgabe konnten so möglicherweise nicht genutzt werden. Außerdem hat bisher, soweit bekannt, noch keine Studie den indirekten Effekt von Feedback über Selbstwirksamkeit auf Kreativität experimentell betrachtet.

4. Eigenständige Konzeption experimenteller Designs zur Analyse des Zusammenhangs von Feedback, Selbstwirksamkeit und Kreativität

Als Weiterentwicklung bisheriger Studien werden nun zwei eigenständig konzipierte experimentelle Designs vorgestellt, die es ermöglichen sollen, den Zusammenhang von Feedback, Selbstwirksamkeit und Kreativität zu veranschaulichen. Anlass für die Entwicklung von gleich zwei Designs geben drei Gründe. Erstens, werden in den experimentellen Designs unterschiedliche Aufgabentypen benutzt, so handelt es sich im ersten Design um eine offene, heuristische Aufgabenstellung, während im zweiten Design eine geschlossene, algorithmische Aufgabe zu lösen ist. Zweitens wird Kreativität unterschiedlich operationalisiert, im ersten Design anhand der CAT und im zweiten Design mittels einer Bewertung der Dimensionen des Divergent Thinking Ansatzes. Drittens, können so unterschiedliche unabhängige Variablen manipuliert werden. Dies sind das Feedbackvorzeichen in Design 1 und der Feedbackstil in Design 2.

4.1. Experimentelles Design mit Feedbackvorzeichen-Treatment

Bei dem nachfolgenden Experimentaldesign handelt es sich um ein between- subjects design, bei dem die Wirkung von positivem oder negativem Feedback auf die Kreativität der VPn beobachtet wird. Weiterhin wird der indirekte Effekt über das psychologische Konzept Selbstwirksamkeit untersucht.

4.1.1. Hypothesen

Die folgenden Hypothesen sollen anhand des experimentellen Designs geprüft werden.

Hypothese 1: Feedback, unabhängig vom Treatment, steigert Kreativität durch die überlieferten Informationen sowie die intrinsische Motivation.

Hypothese 2: Positives Feedback steigert, über den allgemeinen positiven Effekt von Feedback hinaus, die Kreativität der VPn weiterhin, indem es als verbale Überzeugung dient und die kreative Selbstwirksamkeit steigert.

Hypothese 3: Negatives Feedback senkt, im Vergleich zu positiven Feedback, die Kreativität der VPn, indem es die kreativen Fähigkeiten unterminiert und somit die kreative Selbstwirksamkeit senkt.

4.1.2. Aufgabe

Die VPn arbeiten unabhängig voneinander an der in Kapitel 2.1 bereits beschriebenen Postkorb-Fallstudie von Shalley (1991), die häufig in Laborstudien zu Kreativität genutzt wird.¹⁷⁰ Für dieses Experimentaldesign eignen sich Studierende der Wirtschaftswissenschaften, da die Problemstellungen wirtschaftlicher Natur sind und die VPn so ihre im Studium erworbenen Kenntnisse neu kombinieren können, um kreative Lösungen zu erschaffen. Diese Art vorhandenes Wissen zu überdenken, neu zu deuten und umzustrukturieren ist zentral für den kreativen Prozess.¹⁷¹

4.1.3. Verfahren

Das Experiment besteht aus zwei Runden. Bevor jede VP einzeln in den Versuchsraum gebeten wird, wird sie per Zufallseinteilung in eine der drei Versuchsgruppen eingeteilt, die entweder positives, negatives oder kein Feedback erhalten. Danach erhält die VP Instruktionen zu der Aufgabe und die Anweisung, dass kreative, im Sinne von neuartigen und nützlichen Lösungen gefordert sind. Als Nächstes arbeitet die VP an der Aufgabe. Die erste Runde, die acht Memos mit Problemen enthält dauert zehn Minuten. Nach Ablauf der Zeit tritt der/die VersuchsleiterIn in den Raum und teilt der VP mit, dass die Lösungen nun zu einer vierköpfigen Jury gebracht werden, damit diese auf ihre Kreativität hin überprüft werden können. Ferner wird gegenüber der VP erwähnt, dass die Jury aus Personen mit umfangreichen Kenntnissen und Erfahrungen in dem Gebiet besteht. In der Zeit, in der die Lösungen von der Jury ausgewertet werden, sollen die VPn einen Fragebogen über Alter, Geschlecht, Studiengang und Fachsemesterzahl ausfüllen. Personen in der Kontrollgruppe werden nicht über die Jury oder das Feedback informiert, sondern füllen bloß den Fragebogen über ihre demographischen Daten aus. Nach fünf Minuten kehrt der/die VersuchsleiterIn mit dem Lösungsblatt und einem schriftlichen Feedback, je nach entsprechendem Treatment darauf zurück. Es ist insbesondere wichtig, dass das Feedback handschriftlich auf das Lösungsblatt der VP zu schreiben, damit es glaubwürdig und nicht fingiert erscheint.¹⁷² Anschließend erhält die

¹⁷⁰Vgl. Benzer (2006), S. 17; Cooper et al. (1999), S. 45; Shalley (1995), S. 490; Simmons (2011), S. 15; Yuan und Zhou (2008), S. 396.

¹⁷¹Vgl. Guilford (1967a), S. 8.

¹⁷²Vgl. Benzer (2006), S. 63.

¹⁶⁷Vgl. Brockhus et al. (2014), S. 440.

¹⁶⁸Vgl. Pretz und McCollum (2014), S. 230.

¹⁶⁹Vgl. Pretz und McCollum (2014), S. 230.

VP einen Fragebogen mit drei Items zur kreativen Selbstwirksamkeit, die auf einer 5-Punkte Skala von 1 (völlig falsch) bis 5 (völlig richtig) zu bewerten sind. Nach dem Ausfüllen beginnt die zweite Runde und die VP hat eine weitere halbe Stunde Zeit, um die restlichen 14 Memos zu bearbeiten. Danach wird die VP entlassen.

4.1.4. Treatment

Für alle VPn, außer diese der Kontrollgruppe, wird das Feedbackvorzeichen mithilfe der handschriftlichen Feedbackstatements manipuliert. Das Feedback soll hier die zwei Dimensionen verification und elaboration enthalten, also über knowledge of results hinausgehen. Da die VPn durch Zufall in die positive oder negative Versuchsgruppe eingeteilt wurden, aber glauben sollen, dass sie abhängig von ihrer tatsächlichen Leistung Feedback erhalten, ist der elaboration Part fundamental. Deswegen sollen die korrigierenden Informationen Bezug zu der tatsächlichen Leistung haben. Dies geschieht dadurch, dass bestimmte Aspekte genannt werden, die de facto in der Lösung geäußert wurden. Diese Aspekte werden dann relativ zu einem höheren Standard (negatives Feedbackvorzeichen) oder einem niedrigeren Standard (positives Feedbackvorzeichen) bewertet. Das bedeutet mehr Aufwand für den/die VersuchsleiterIn, welche/r sich jede Aufgabenlösung schon während des Experimentablaufs anschauen muss, erhöht jedoch die Glaubwürdigkeit des Feedbacks erheblich.¹⁷³ Dementsprechend erhalten VPn in der positiven Feedback Versuchsgruppe das folgende Feedbackstatement inklusive einer Anmerkung warum die Leistung lobenswert war:

„Herzlichen Glückwunsch, du hast im Vergleich zum Durchschnitt sehr kreative Lösungen für die vorgestellten Probleme gefunden. Deine Lösungen lassen auf deine Fähigkeit zur kreativen Problemlösung schließen. Vor allem, dass du nicht nur versucht hast Maßnahmen im Betriebskontext durchzusetzen, sondern vielmehr das private Umfeld der MitarbeiterInnen einbezogen hast, hat unsere Jury überzeugt.“

Analog erhalten VPn in der negativen Feedback Versuchsgruppe ein Feedback mit Begründung und korrigierenden Informationen für die nächste Runde:

„Leider hast du, relativ zum Durchschnitt, kaum kreative Lösungen für die vorgestellten Probleme gefunden. Deine Lösungen waren leider alle sehr konventionell und wenig originell. Du hast selten versucht über eine normale, gebräuchliche Idee hinauszugehen. Vielleicht hilft es dir, wenn du bei der nächsten Runde versuchst, nicht nur über Maßnahmen im Betrieb nachzudenken, sondern dir vielmehr das private Umfeld der MitarbeiterInnen anschaut.“

Die elaboration Bestandteile der Feedbackstatements sind dabei variabel und sollten auf die tatsächliche Leistung angepasst werden.

4.1.5. Messgrößen

Da es ein bewährtes Verfahren ist, eine vorhandene Skala mit akzeptablen psychometrischen Gütekriterien zu nutzen¹⁷⁴, wird in diesem Experiment kreative Selbstwirksamkeit anhand der 3 Items von Beghetto (2006) operationalisiert und gemessen. Diese sind erstens „Ich bin gut darin, mir neue Ideen einfallen zu lassen“, zweitens „Ich habe viele gute Ideen“ und drittens „Ich habe ein gutes Vorstellungsvermögen“.¹⁷⁵ Wie in Kapitel 3.2 beschrieben und in Tabelle 2 im Anhang dargestellt, werden die Items von Beghetto (2006) identisch oder adaptiert in Studien häufig genutzt. Da die Items alle positiv formuliert sind, ist es möglich durch Addition einen Score für die kreative Selbstwirksamkeit für jede VP zu errechnen.

Bei der Bewertung von Kreativität wird sich an früheren Studien orientiert, die ebenfalls die Aufgabe von Shalley (1991) anwenden, und die kreative Performance mithilfe der CAT von Amabile (1982) beurteilen.¹⁷⁶ Daraus ergeben sich gewisse Bedingungen für die Jury, die in Kapitel 2.1 beschrieben wurden, und die nachfolgenden Implikationen haben. Bezüglich der Angemessenheit der RichterInnen lässt sich feststellen, dass es sich bei den 22 Memos von Shalley (1991) um Problemstellungen aus dem Personalwesen handelt, weshalb bei den bisherigen Studien teilweise HochschulabsolventInnen mit langjähriger Berufserfahrung im Personalbereich als RichterInnen engagiert wurden.¹⁷⁷ Hier sollen anteilig Personen mit Erfahrung im Personalbereich und zusätzlich Studierende, die im Rahmen ihres Wirtschaftswissenschaftsstudium einen Einblick in die Materie haben, in der Jury sitzen. Die unabhängige Bewertung der Jury soll durch räumliche Distanz und der bewussten Diskretion über Kreativitätskriterien erreicht werden. Hinsichtlich der Anforderung des relativen Bewertungsschemas soll, nach dem Beispiel von Zhou (1998), bei dem erarbeiteten Experimentaldesign der vorliegenden von dem/der VersuchsleiterIn angewiesen werden, dass die RichterInnen vor der eigentlichen Bewertungsphase zehn zufällig ausgewählte Lösungen zu jedem Memo zu lesen, damit diese eine Vorstellung von dem Spektrum an Kreativität der VPn bekommen und dementsprechend relativ bewerten.¹⁷⁸ Angesichts der möglichen Reihenfolge-Effekte, sollten die Lösungen den RichterInnen jeweils in unterschiedlicher Aufeinanderfolge gezeigt werden. Im Hinblick auf die Anzahl der RichterInnen werden hier unter Berücksichtigung der anderen Studien, die eine Anzahl zwischen zwei und

¹⁷⁴Vgl. Puente-Díaz (2016), S. 177f.

¹⁷⁵Vgl. Beghetto (2006), S. 450.

¹⁷⁶Vgl. Benzer (2006), S. 19f.; Cooper et al. (1999), S. 47f.; Shalley (1991), S. 182; Shalley (1995), S. 491; Simmons (2011), S. 15; Yuan und Zhou (2008), S. 398f.; Zhou (1998), S. 268.

¹⁷⁷Vgl. Shalley (1991), S. 182; Yuan und Zhou (2008), S. 398.

¹⁷⁸Vgl. Zhou (1998), S. 268.

¹⁷³Vgl. Benzer (2006), S. 63.

sechs gewählt haben¹⁷⁹, vier RichterInnen in der Jury sein. Die vier RichterInnen sollen die Kreativität der Antworten auf einer 6-Punkte Skala von 1 (= nicht kreativ) bis 6 (= sehr kreativ) bewerten, um eine Überforderung durch zu viele Abstufungen zu vermeiden.¹⁸⁰

4.2. Experimentelles Design mit Feedbackstil-Treatment

Das folgende Experimentaldesign, ebenfalls ein between-subjects design macht die Wirkung eines informativen oder kontrollierenden Feedbackstils auf die Kreativität der VPn beobachtbar. Das psychologische Konzept der Selbstwirksamkeit wird außen vorgelassen, da literaturbasiert kein Zusammenhang zwischen dem Feedbackstil und der Selbstwirksamkeit festgestellt werden konnte.

4.2.1. Hypothesen

Die folgenden Hypothesen sollen anhand des experimentellen Designs geprüft werden.

Hypothese 1: Feedback, unabhängig vom Treatment, steigert Kreativität durch die überlieferten Informationen.

Hypothese 2: Informatives Feedback steigert, über den allgemeinen positiven Effekt von Feedback hinaus, die Kreativität der VPn weiterhin, indem es die intrinsische Motivation der VPn durch Fokus auf die Aufgabe steigert.

Hypothese 3: Kontrollierendes Feedback senkt, im Vergleich zum informativen Feedback, die Kreativität der VPn, indem es den Fokus auf externe Zwänge legt und so die intrinsische Motivation unterminiert.

4.2.2. Aufgabe

Die VPn arbeiten unabhängig voneinander an der algorithmischen Unusual Uses Aufgabe aus dem TTCT, die bereits in Kapitel 2.1 aufgegriffen wurde. Trotz Amabiles (1983) Einschätzung, dass kreative Leistung nur bei heuristischen Aufgaben möglich ist, wird diese Aufgabe oftmals zur Beobachtung von kreativen Fähigkeiten eingesetzt¹⁸¹ und erhält dadurch ihre Legitimation für die vorliegende Arbeit. Gemäß den Vorgaben von Torrance (1974a) haben die VPn jeweils zehn Minuten Zeit, sich ungewöhnliche Zwecke für die Objekte Pappkarton und Konservendose auszudenken.¹⁸² Im Gegensatz zur Postkorb- Fallstudie in ersten Experimentaldesign ist diese Aufgabe für alle möglichen Personen geeignet, da der TTCT als Kreativitätstest für Personen vom Kindergarten bis ins Erwachsenenalter gedacht ist.¹⁸³

¹⁷⁹Vgl. Benzer (2006), S. 24; Cooper et al. (1999), S. 47; Shalley (1991), S. 128; Shalley (1995), S. 491; Simmons (2011), S. 15; Yuan und Zhou (2008), S. 398; Zhou (1998), S. 268.

¹⁸⁰Vgl. Raab-Steiner und Benesch (2015), S. 60.

¹⁸¹Vgl. Batey et al. (2010), S. 533; Bradler et al. (2016), S. 6f.; Furnham et al. (2008), S. 1064.

¹⁸²Vgl. Torrance (1974a), S. 7.

¹⁸³Vgl. Torrance (1974b), S. 4.

4.2.3. Verfahren

Dieses Experimentaldesign ähnelt im Verlauf dem Experimentaldesign 1, denn es besteht ebenso aus zwei Runden. Bevor jede VP einzeln in den Versuchsraum gebeten wird, wird sie ebenfalls per Zufallseinteilung in eine der drei Versuchsgruppen eingeteilt, die in diesem Fall informatives, kontrollierendes oder kein Feedback umfassen. Danach erhält sie die von Torrance (1974a) ausformulierten Instruktionen zu der Aufgabe.¹⁸⁴ Als Nächstes arbeitet die VP an der Aufgabe. In der ersten Runde sollen in zehn Minuten Unusual Uses für das Objekt Pappkarton gefunden werden. Nach Ablauf der Zeit tritt der/die VersuchsleiterIn in den Raum und teilt der VP mit, dass die Lösungen nun zu einer vierköpfigen Jury gebracht werden, damit diese auf ihre Kreativität hin überprüft werden können. Außerdem wird gegenüber der VP erwähnt, dass die Jury retrospektive Vergleichswerte besitzt, anhand derer sie die relative Kreativität bewerten können. In der Zeit, in der Lösungen von der Jury ausgewertet werden, sollen die VPn einen Fragebogen über ihr Alter, Geschlecht und Tätigkeit ausfüllen. Personen in der Kontrollgruppe werden nicht über die Jury oder das Feedback informiert, sondern füllen bloß den Fragebogen über ihre demographischen Daten aus. Nach fünf Minuten kehrt der/die VersuchsleiterIn mit dem Lösungsblatt und einem schriftlichen Feedback je nach entsprechendem Treatment darauf zurück. Wie in Experiment 1 ist es hier um der Glaubwürdigkeit willen, wichtig, das Feedback handschriftlich auf das Lösungsblatt der VP zu schreiben. Nachdem das Feedback der VP überbracht wurde, bricht die zweite Runde an und die VP hat wieder zehn Minuten Zeit um Unusual Uses für das zweite Objekt, die Konservendose, zu entwickeln. Danach wird die VP entlassen.

4.2.4. Treatment

Für alle VPn, außer die der Kontrollgruppe, wird der Feedbackstil mithilfe der handschriftlichen Feedbackstatements manipuliert. Es ist hier ebenso wie in Experimentaldesign 1 essentiell, dass das Feedback die zwei Dimensionen verification und elaboration enthält. Hier liegt der Fokus jedoch nicht auf der Glaubwürdigkeit, die durch den elaboration Part generiert werden soll, sondern vielmehr auf den Informationen zu den Anforderungen der Aufgabe. In Anlehnung an Zhou (1998) wird der kontrollierende Feedbackstil mit einem you should Statement artikuliert. Der informative Feedbackstil bleibt neutral und liefert lediglich ein Lob für gute Leistung oder eine Ermahnung bei schlechter Leistung. Unabhängig vom Treatment, wird im elaboration Part des Feedbacks, Bezug darauf genommen, was die VP gut oder schlecht gemacht hat. Dies lässt sich bei der Unusual Uses Aufgabe einfach lösen, da ohnehin auf drei Dimensionen bewertet wird und der Scoring Guide dem/der ExperimentatorIn eine schnelle Bewertung der Aufgaben ermöglicht. Dementsprechend erhalten die VPn in dem kontrollierenden Feedback-Treatment bei guter Leistung das folgende Statement nebst variabler Anmerkung, warum ihre Leistung anerkennenswert ist:

¹⁸⁴Siehe Abbildung 6, Anhang.

Herzlichen Glückwunsch, du hast die Aufgabe, genau wie du solltest, kreativ gelöst. Deine Antworten lassen auf deine Fähigkeit zur kreativen Problemlösung schließen. Vor allem, dass du nicht nur originelle, d.h. statistisch selten genannte, sondern auch unterschiedlichen Kategorien angehörende Verwendungszwecke gefunden hast, hat unsere Jury überzeugt. Pass auf, dass du in der nächsten Runde ebenfalls kreative Verwendungszwecke findest, ansonsten können wir deine Daten nicht verwenden.“.

Entsprechend wird bei schlechter Leistung das folgende informative Statement mitgeteilt sowie im variablen Teil angemerkt, worin Fehler bestanden haben und was besser gemacht werden könnte:

„Du hast bei der Aufgabe schlecht abgeschnitten und keine kreativen Antworten geliefert. Deine Antworten haben wenig variiert, sondern gehörten alle in eine Kategorie. Du solltest bei der nächsten Runde versuchen, unterschiedliche Funktionen des Objekts zu beachten und dich nicht nur auf eine Funktion beschränken, wie beispielsweise einen Pappkarton als Behältnis zu nutzen. Wenn du das nicht tust und sich die schlechte Leistung wiederholt, können wir deine Daten nicht verwenden“.

Die zweiten Teile der Feedbackstatements sind dabei variabel und sollten auf die tatsächliche Leistung angepasst werden.

4.2.5. Messgrößen

Die kreative Performance wird gemäß den Instruktionen von *Torrance (1974a)* auf den in Kapitel 2.1 beschriebenen Dimensionen fluency, flexibility und originality ausgewertet.¹⁸⁵

5. Fazit

In dieser Arbeit kann keine endgültige Antwort auf die Frage gegeben werden, unter welchen Bedingungen Feedback einen Effekt auf die Kreativität von Personen hat. Dies liegt vor allem an der Komplexität menschlicher Verhaltensweisen, sowie der fortwährenden Unklarheit über die kreativen Prozesse. Trotzdem lassen sich Empfehlungen gegenüber Führungskräften geben, die das Instrument Feedback nutzen wollen, um die Kreativität ihrer Angestellten zu fördern.

Zuerst, sollte ein Bewusstsein darüber entstehen, dass Kreativität von Individuen zum Großteil auf stabilen Persönlichkeitsfaktoren, wie Risikofreude und Non-Konformität beruht, welche durch Feedback schwerlich verändert werden können. Der Bestandteil von Kreativität, an dem Feedback anknüpfen kann, ist die aufgabenspezifische Motivation. Damit die Gründe für die Durchführung einer Tätigkeit durch

ein intrinsisches Interesse geprägt sind, sollte das Individuum frei von Kontrolle und ohne externen Druck arbeiten. Im Hinblick auf diese Einschränkung lässt sich die Bedeutsamkeit von informativen und die Gefahr von kontrollierendem Feedback erkennen. Kontrollierendes Feedback verschiebt den Fokus von der Aufgabe hin zu einem externen Ziel und macht dem/der Feedback-EmpfängerIn bewusst, dass er/sie nicht unabhängig und selbstbestimmt handelt. Dieser Fokus hemmt Personen meist bei der Produktion von kreativen Ideen. Informatives Feedback hingegen stellt den/die Feedback-EmpfängerIn in das Zentrum und vermittelt lediglich Informationen und Hinweise. So wird die Fokussierung auf die Aufgabe gewahrt und eine intrinsische Aufgabenmotivation stellt sich ein, wenn eine grundsätzlich positive Einstellung zu der kreativen Aufgabe vorhanden ist. Ein informativer Feedbackstil lässt sich zudem unabhängig von der Qualität der Leistung anwenden, so kann und sollte auch negatives Feedback in einem informativen Stil gegeben werden, um intrinsisches Interesse an der Aufgabe zu bewahren. Um die Selbstwirksamkeit der Angestellten, zu steigern, kann jedoch nur positives Feedback angewandt werden, da allein dieses als verbale Überzeugung fungiert. Schlussendlich ist Feedback, losgelöst von Feedbackstil und – vorzeichen, stets zu empfehlen, um Informationen über Anforderungen und Ziele der kreativen Aufgabe zu vermitteln.¹⁸⁶

Bei alledem sollte Feedback immer genügend Informationen enthalten und über reines knowledge of results hinausgehen, damit eine Wirkung erzielt werden kann. Hinsichtlich der Bewertung der Kreativität, welche als Information in das Feedback eingeht, lassen sich Empfehlungen aus der CAT und dem TTCT gewinnen. So ist eine Folgerung, die aus der CAT gezogen werden kann, dass Kreativität nicht absolut, sondern immer im Vergleich zu vergangenen Leistungen oder entsprechenden Personen zu bewerten ist. Auf diese Weise sollte bei Personen, die zwar relativ zu anderen Personen weniger kreativ, aber relativ zu vergangenen Leistungen sehr kreativ waren, die Steigerung der Kreativität einbezogen werden.¹⁸⁷ Daneben veranschaulicht der TTCT, dass verschiedene Dimensionen der Kreativität bewertet werden können.

Zuletzt, lässt sich sagen, dass die Ergebnisse der Arbeit als Grundlage für experimentelle Forschung zum Thema Feedback und Kreativität dienen kann. Jedoch sind hierzu Experimente mit genügend Probanden notwendig. Zugleich ist die größte Limitation der Arbeit dementsprechend, dass die experimentellen Designs nur beschrieben und nicht ausgeführt wurden.

¹⁸⁶Überblick über die positive Wirkung von Feedback auf Kreativität in Abbildung 6, Anhang.

¹⁸⁷Vgl. *Zhou (1998)*, S. 273.

¹⁸⁵Vgl. *Torrance (1974a)*, S. 31.

Literatur

- Amabile, T. M. *Componential theory of creativity*. Kessler, E. H. (Hrsg.): Encyclopedia of Management Theory, Thousand Oaks, S. 134-139, 2013.
- Amabile, T. und Gryskiewicz, S. S. *Creativity in the R&D laboratory*. Center for Creative Leadership, 1987.
- Amabile, T. M. Social psychology of creativity: A consensual assessment technique. *Journal of personality and social psychology*, 43(5):997-1013, 1982.
- Amabile, T. M. The social psychology of creativity: A componential conceptualization. *Journal of personality and social psychology*, 45(2):357-376, 1983a.
- Amabile, T. M. *Creativity in context: Update to the social psychology of creativity*. Hachette UK, 1996.
- Amabile, T. The social psychology of creativity: A componential conceptualization. 1983b.
- Ammons, R. B. Effects of knowledge of performance: A survey and tentative theoretical formulation. *The Journal of general psychology*, 54(2):279-299, 1956.
- Ashford, S. J. und Cummings, L. L. Feedback as an individual resource: Personal strategies of creating information. *Organizational behavior and human performance*, 32(3):370-398, 1983.
- Baer, M., Oldham, G. R., und Cummings, A. Rewarding creativity: when does it really matter? *The Leadership Quarterly*, 14(4-5):569-586, 2003.
- Bandura, A. Self-efficacy. in: Friedman, H. S. (Hrsg.): *Encyclopedia of Mental Health, San Diego*, pages 421-432, 1998.
- Bandura, A. Self-efficacy: toward a unifying theory of behavioral change. *Psychological review*, 84(2):191-215, 1977.
- Bandura, A. Self-efficacy: The exercise of control, 1. Aufl., New York. pages 1-35, 1997.
- Barron, F. und Harrington, D. M. Creativity, intelligence, and personality. *Annual review of psychology*, 32(1):439-476, 1981.
- Barysch, K. N. Selbstwirksamkeit. In *Psychologie der Werte*, pages 201-211. Springer, 2016.
- Batey, M., Furnham, A., und Safiullina, X. Intelligence, general knowledge and personality as predictors of creativity. *Learning and individual differences*, 20(5):532-535, 2010.
- Beckers, N., Cools, M., Van den Abbeele, A., et al. The impact of incentives on creativity: A literature review. *Review of Business and Economic Literature*, 55(4):463-484, 2010.
- Beghetto, R. A. Creative self-efficacy: Correlates in middle and secondary students. *Creativity Research Journal*, 18(4):447-457, 2006.
- Beghetto, R. A., Kaufman, J. C., und Baxter, J. Answering the unexpected questions: Exploring the relationship between students' creative self-efficacy and teacher ratings of creativity. *Psychology of Aesthetics, Creativity, and the Arts*, 5(4):342-349, 2011.
- Benzer, J. *You must be Creative! The Effect of Performance Feedback on Intrinsic Motivation and Creativity, Dissertation, Texas A&M University*. PhD thesis, 2006.
- Bradler, C., Neckermann, S., und Warnke, A. *Incentivizing Creativity: A Large-Scale Experiment with Tournaments and Gifts, Arbeitspapier, Zentrum für Europäische Wirtschaftsforschung GmbH*. 2016.
- Brockhus, S., Van der Kolk, T., Koeman, B., und Badke-Schaub, P. The influence of creative self-efficacy on creative performance. In *Proceedings of the DESIGN 2014 13th International Design Conference*, pages 437-444, 2014.
- Bruno, F. J. *Psychology: A self-teaching guide*, volume 182. John Wiley & Sons, 2002.
- Byron, K. und Khazanchi, S. Rewards and creative performance: a meta-analytic test of theoretically derived hypotheses. *Psychological bulletin*, 138(4):809-830, 2012.
- Carson, P. P. und Carson, K. D. Managing creativity enhancement through goal-setting and feedback. *The Journal of Creative Behavior*, 27(1):36-45, 1993.
- CASAS-ARCE, P., Lourenço, S. M., und MARTÍNEZ-JEREZ, F. A. The performance effect of feedback frequency and detail: Evidence from a field experiment in customer satisfaction. *Journal of Accounting Research*, 55(5):1051-1088, 2017.
- Cervone, D., Jiwani, N., und Wood, R. Goal setting and the differential influence of self-regulatory processes on complex decision-making performance. *Journal of personality and social psychology*, 61(2):257-266, 1991.
- Chan, J. Researching creativity and creativity research. *Handbook of research on creativity*, pages 21-32, 2013.
- Charness, G. und Grieco, D. Creativity and financial incentives. *University of California, Santa Barbara, Working Paper*, 2014.
- Cooper, B. L., Clasen, P., Silva-Jalonen, D. E., und Butler, M. C. Creative performance on an in-basket exercise: Effects of inoculation against extrinsic reward. *Journal of Managerial Psychology*, 14(1):39-57, 1999.
- Cortina, J. M. What is coefficient alpha? an examination of theory and applications. *Journal of applied psychology*, 78(1):98-104, 1993.
- Cropley, D. H. und Kaufman, J. C. 14. rating the creativity of products. *Handbook of research on creativity*, pages 196-211, 2013.
- Crutchfield, R. Conformity and creative thinking. In in: Gruber, H. / Terrell, G. / Wertheimer, M. (Hrsg.): *Contemporary Approaches to Creative Thinking: A Symposium held at the University of Colorado, New York*, pages 120-140. Atherton Press, 1962.
- Deci, E. L. *Intrinsic Motivation*. 1. Aufl., New York, 1975.
- Deci, E. und Ryan, R. M. *Intrinsic motivation and self-determination in human behavior*. Springer Science & Business Media, 1985.
- Diehl, M. und Munkes, J. Kreativität und innovation. *Theorien der Sozialpsychologie*, 2:366-389, 2002.
- Dimotakis, N., Mitchell, D., und Maurer, T. Positive and negative assessment center feedback in relation to development self-efficacy, feedback seeking, and promotion. *Journal of Applied Psychology*, 102(11):1514-1527, 2017.
- Dörflinger, S. Kreativität - unternehmen sollten ihren mitarbeitern mehr zutrauen, 2016. URL <https://www.zeit.de/karriere/beruf/2016-08/kreativitaet-innovation-brainstorming-teamarbeit-fuehrungskraft>. abgerufen am: 03.03.2018.
- Drever, J. und Fröhlich, W. D. *Wörterbuch zur Psychologie*. 1. Aufl., München, 1972.
- Eisenberger, R., Armeli, S., und Pretz, J. Can the promise of reward increase creativity? *Journal of Personality and Social Psychology*, 74(3):704-714, 1998.
- Fengler, J. Feedback als interventions-methode. *Gruppendynamik und Organisationsberatung*, 41(1):5-20, 2010.
- Fodor, E. M. und Carver, R. A. Achievement and power motives, performance feedback, and creativity. *Journal of Research in Personality*, 34(4):380-396, 2000.
- Fombelle, P. W., Bone, S. A., und Lemon, K. N. Responding to the 98%: face-enhancing strategies for dealing with rejected customer ideas. *Journal of the Academy of Marketing Science*, 44(6):685-706, 2016.
- Friedman, R. S. Reinvestigating the effects of promised reward on creativity. *Creativity Research Journal*, 21(2-3):258-264, 2009.
- Furnham, A., Batey, M., Anand, K., und Manfield, J. Personality, hypomania, intelligence and creativity. *Personality and Individual Differences*, 44(5):1060-1069, 2008.
- George, D. und Mallery, M. Using spss for windows step by step: a simple guide and reference. 2003.
- Gill, R. The in-tray (in-basket) exercise as a measure of management potential. *Journal of Occupational Psychology*, 52(3):185-197, 1979.
- Gross, D. Creativity under fire: The effects of competition on creative production. *Working Paper Harvard Business School*, 2016.
- Guilford, J. P. Creativity: Yesterday, today and tomorrow. *The Journal of Creative Behavior*, 1(1):3-14, 1967a.
- Guilford, J. P. The nature of human intelligence. 1967b.
- Hass, R. W., Katz-Buonincontro, J., und Reiter-Palmon, R. Disentangling creative mindsets from creative self-efficacy and creative identity: Do people hold fixed and growth theories of creativity? *Psychology of Aesthetics, Creativity, and the Arts*, 10(4):436-446, 2016.
- Hattie, J. und Timperley, H. The power of feedback. *Review of educational research*, 77(1):81-112, 2007.
- Ilgel, D. R., Fisher, C. D., und Taylor, M. S. Consequences of individual feedback on behavior in organizations. *Journal of applied psychology*, 64(4):349-371, 1979.
- Jaussi, K. S., Randel, A. E., und Dionne, S. D. I am, i think i can, and i do: The role of personal identity, self-efficacy, and cross-application of experiences in creativity at work. *Creativity Research Journal*, 19(2-3):247-258, 2007.
- Karwowski, M. Did curiosity kill the cat? relationship between trait curiosity, creative self-efficacy and creative personal identity. *Europe's Journal of Psychology*, 8(4):547-558, 2012.

- Karwowski, M., Lebuda, I., und Wiśniewska, E. Measurement of creative self-efficacy and creative role-identity. *High Ability Studies*, 22:291–231, 2012.
- Kluger, A. N. und DeNisi, A. The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological bulletin*, 119(2):254–284, 1996.
- Kozbelt, A., Beghetto, R. A., und Runco, M. A. Theories of creativity. *The Cambridge handbook of creativity*, 2:20–47, 2010.
- Kulhavy, R. W. Feedback in written instruction. *Review of educational research*, 47(2):211–232, 1977.
- Kulhavy, R. W. und Stock, W. A. Feedback in written instruction: The place of response certainty. *Educational Psychology Review*, 1(4):279–308, 1989.
- Latham, G. P. und Locke, E. A. Self-regulation through goal setting. *Organizational behavior and human decision processes*, 50(2):212–247, 1991.
- Lemons, G. Bar drinks, rugas, and gay pride parades: Is creative behavior a function of creative self-efficacy? *Creativity Research Journal*, 22(2): 151–161, 2010.
- Luthans, F., Youssef, C. M., und Avolio, B. J. *Psychological capital: Developing the human competitive edge*, volume 198. Oxford University Press Oxford, 2007.
- Mathisen, G. E. und Bronnick, K. S. Creative self-efficacy: An intervention study. *International Journal of Educational Research*, 48(1):21–29, 2009.
- Otley, D. Performance management: a framework for management control systems research. *Management accounting research*, 10(4):363–382, 1999.
- Pretz, J. E. und McCollum, V. A. Self-perceptions of creativity do not always reflect actual creative performance. *Psychology of Aesthetics, Creativity, and the Arts*, 8(2):227–236, 2014.
- Puente-Díaz, R. Creative self-efficacy: An exploration of its antecedents, consequences, and applied implications. *The Journal of psychology*, 150(2): 173–193, 2016.
- Puente-Díaz, R. und Cavazos-Arroyo, J. The influence of creative mindsets on achievement goals, enjoyment, creative self-efficacy and performance among business students. *Thinking Skills and Creativity*, 24:1–11, 2017.
- Qian, J., Yang, F., und Han, Z. R. The influencing mechanisms of authentic leadership on feedback-seeking behaviour: a cognitive/emotive model. *Current Psychology*, 35(3):478–485, 2016.
- Quarch, C. Die zukunft der arbeit - das büro als kreative keimzelle, 2018. URL https://www.deutschlandfunk.de/die-zukunft-der-arbeit-das-buero-als-kreative-keimzelle.807.de.html?dram:article_id=407413. abgerufen am: 03.03.2018.
- Raab-Steiner, E. und Benesch, M. *Der Fragebogen*. 4. Aufl., Wien, 2015.
- Ramaprasad, A. On the definition of feedback. *Behavioral Science*, 28(1): 4–13, 1983.
- Reber, A. S. und Reber, E. S. *The Penguin Dictionary of Psychology*. 3. Aufl., London, 2001.
- Reiter-Palmon, R., Robinson-Morrall, E. J., Kaufman, J. C., und Santo, J. B. Evaluation of self-perceptions of creativity: Is it a useful criterion? *Creativity Research Journal*, 24(2-3):107–114, 2012.
- Richter, A. W., Hirst, G., Van Knippenberg, D., und Baer, M. Creative self-efficacy and individual creativity in team contexts: Cross-level interactions with team informational resources. *Journal of Applied Psychology*, 97(6):1282–1290, 2012.
- Rodgers, W. M., Markland, D., Selzler, A.-M., Murray, T. C., und Wilson, P. M. Distinguishing perceived competence and self-efficacy: An example from exercise. *Research quarterly for exercise and sport*, 85(4):527–539, 2014.
- Ryan, R. M. Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of personality and social psychology*, 43(3):450–461, 1982.
- Schwarzer, R. und Jerusalem, M. Das Konzept der Selbstwirksamkeit, in: Jerusalem, M. / Hopf, D. (Hrsg.): *Selbstwirksamkeit und Motivationsprozesse in Bildungsinstitutionen*, Weinheim, S. 28–35, 2002.
- Shalley, C. E. Effects of productivity goals, creativity goals, and personal discretion on individual creativity. *Journal of Applied psychology*, 76(2): 179–185, 1991.
- Shalley, C. E. Effects of coaction, expected evaluation, and goal setting on creativity and productivity. *Academy of Management Journal*, 38(2):483–503, 1995.
- Shalley, C. E. und Gilson, L. L. What leaders need to know: A review of social and contextual factors that can foster or hinder creativity. *The leadership quarterly*, 15(1):33–53, 2004.
- Shute, V. J. Focus on formative feedback. *Review of educational research*, 78(1):153–189, 2008.
- Simmons, A. L. The influence of openness to experience and organizational justice on creativity. *Creativity Research Journal*, 23(1):9–23, 2011.
- Tierney, P. und Farmer, S. M. Creative self-efficacy: Its potential antecedents and relationship to creative performance. *Academy of Management journal*, 45(6):1137–1148, 2002.
- Tierney, P. und Farmer, S. M. Creative self-efficacy development and creative performance over time. *Journal of Applied Psychology*, 96(2):277–293, 2011.
- Tierney, P., Farmer, S. M., und Graen, G. B. An examination of leadership and employee creativity: The relevance of traits and relationships. *Personnel psychology*, 52(3):591–620, 1999.
- Torrance, E. P. *Torrance Tests of Creative Thinking: Directions Manual and Scoring Guide*, Lexington, 1974a.
- Torrance, E. P. *Torrance Tests of Creative Thinking: Norms-Technical Manual*, Lexington, 1974b.
- Utman, C. H. Performance effects of motivational state: A meta-analysis. *Personality and Social Psychology Review*, 1(2):170–182, 1997.
- Wang, X., Schneider, C., und Valacich, J. S. Enhancing creativity in group collaboration: How performance targets and feedback shape perceptions and idea generation performance. *Computers in Human Behavior*, 42: 187–195, 2015.
- Wirtz, M. A. *Dorsch-Lexikon der psychologie*. 18. Aufl., Bern, 2014.
- Yuan, F. und Zhou, J. Differential effects of expected external evaluation on different parts of the creative idea production process and on final product creativity. *Creativity Research Journal*, 20(4):391–403, 2008.
- Zhou, J. Feedback valence, feedback style, task autonomy, and achievement orientation: Interactive effects on creative performance. *Journal of applied psychology*, 83(2):261–276, 1998.
- Zhou, J. Promoting creativity through feedback. *Handbook of organizational creativity*, pages 125–145, 2008.
- Zhou, J. und George, J. M. When job dissatisfaction leads to creativity: Encouraging the expression of voice. *Academy of Management journal*, 44(4):682–696, 2001.
- Zhou, J. und Shalley, C. E. Research on employee creativity: A critical review and directions for future research. In *Research in personnel and human resources management*, pages 165–217. Emerald Group Publishing Limited, 2003.



Besteuerung unternehmensverbundener Stiftungen

Iris Krampe

Westfälische Hochschule Gelsenkirchen

Abstract

4,3 Mrd. Euro jährliche Stiftungsausgaben für Satzungszwecke. Das entspricht der Höhe des Risikokapitals, das 2017 in Start-Ups in Deutschland investiert wurde.“, titulierte der Bundesverband deutscher Stiftungen. Das zeigt die eindrucksvolle Dimension der Stiftungslandschaft in Deutschland. Die Motivation zur Gründung einer Stiftung kann vielfältige Facetten haben. Meist ist nicht eine einzelne Zielsetzung ausschlaggebend eine Stiftung ins Leben zu rufen. Unternehmensverbundene Stiftungen zeichnen sich dadurch aus, dass ein (enger) Bezug zu einem Unternehmen besteht. So haben einige Unternehmer durch eine Stiftung die Möglichkeit geschaffen, ihr Lebenswerk auf Dauer zu erhalten. Zu den wohl bekanntesten unternehmensverbundenen Stiftungen zählen beispielsweise die Alfried Krupp von Bohlen und Halbach-Stiftung, die Fritz Thyssen Stiftung oder die Robert Bosch Stiftung. Bei Gründung einer rechtsfähigen Stiftung ergeben sich für den Stifter Gestaltungsspielräume beim Stiftungszweck, Stiftungsvermögen und bei der Stiftungsorganisation. Das Hauptaugenmerk dieser Arbeit liegt in der Analyse der Besteuerung der Stiftungen. Es wird dargelegt, inwiefern die Stiftung bei Errichtung, während des Bestehens und bei Auflösung besteuert wird. Zudem liegt der Fokus dieser Arbeit auf den möglichen Steuerbegünstigungen und den notwendigen Voraussetzungen hierfür, da 95 % aller Stiftungen gemeinnützige Zwecke verfolgen. Zuletzt kommt als besondere Gestaltungsform für Unternehmer die Kombination einer privatnützigen Familienstiftung mit einer gemeinnützigen Stiftung zur sogenannten Doppelstiftung in Betracht.

Keywords: Unternehmensstiftung, Unternehmensnachfolge, Gemeinnützigkeit, Steuerbegünstigungen

1. Einleitung

Bei der Planung der Unternehmensnachfolge spielen verschiedene Stiftungsgestaltungen eine immer bedeutsamere Rolle. Die Stiftung wird dabei als vorteilhafte Form zur Sicherung der Unternehmenskontinuität angesehen und kann als Mittel zum Schutz vor Zerschlagung des vom Stifter aufgebauten Lebenswerkes dienen. Auch viele bekannte deutsche Unternehmen weisen heute einen direkten oder indirekten Stiftungsbezug auf. So sind Aldi, Adidas, Thyssen, Vorwerk und Grundig nur einige Beispiele von Unternehmen mit Stiftungsbezug.¹ Insgesamt hat jede zehnte rechtsfähige Stiftung einen Unternehmenshintergrund.²

Im Rahmen dieser Arbeit sollen die Besonderheiten der unternehmensverbundenen Stiftung aufgezeigt werden. Es wird dabei genauer auf die Chancen und Risiken im rechtlichen und steuerrechtlichen Bereich eingegangen. Die Arbeit

soll außerdem die Erstellung eines geeigneten Stiftungskonzeptes ermöglichen, indem die kritischen Faktoren hierfür erläutert werden. Sie soll als Entscheidungshilfe für potentielle Stifter dienen, um herauszufinden, ob eine Stiftungslösung als Nachfolgeregelung in Frage kommt.

Beim Begriff der Stiftung handelt es sich im Folgenden, falls nicht genauer gekennzeichnet, um die rechtsfähige Stiftung bürgerlichen Rechts. Der Fokus soll dabei insbesondere auf den unternehmensverbundenen Stiftungen liegen. Auf die nicht unwichtigen weiteren Gestaltungsalternativen zur rechtsfähigen Stiftung wie z. B. die unselbstständige Stiftung (sog. treuhänderische Stiftung), die Stiftungs-GmbH oder den Stiftungsverein soll hier nicht genauer eingegangen werden.

Zu Beginn dieser Arbeit werden die Grundlagen des Stiftungswesens erläutert. Es soll die Frage beantwortet werden, welche Gründe einen vermögenden Stifter bzw. Unternehmer dazu veranlassen können, sein Vermögen auf eine Stiftung zu übertragen. Dabei spielen nicht nur ideelle Motive eine Rolle, es können auch wirtschaftliche Beweggründe, wie

¹Vgl. G. and T. (2014), S. 3

²Vgl. Götz in Götz and PachHanssenheimb (2016), Rz. 3

die Fortführung des Unternehmens, relevant sein. Des Weiteren wird erläutert, wie eine rechtsfähige Stiftung gegründet wird und welche Voraussetzungen für eine Anerkennung notwendig sind. Hier wird genauer auf die Anforderungen an Stiftungszweck, Stiftungsvermögen und die Stiftungsorganisation als zentrale Gestaltungspunkte des Stiftungsaufbaus eingegangen. Zu den grundlegenden Aspekten einer Stiftung zählt auch die staatliche Überwachung durch die Stiftungsaufsicht. Hier werden die Aufgaben der Stiftungsaufsicht und die Begrenzung ihrer Tätigkeitsbereiche aufgegriffen.

Im Rahmen dieser Arbeit wird dann zunächst die Besteuerung der privatnützigen unternehmensverbundenen Stiftung erläutert. Diese variiert je nach Tatbestand, sie unterscheidet sich also bei Errichtung der Stiftung, während der laufenden Tätigkeit und bei Stiftungsauflösung. Grundsätzlich ähnelt die Besteuerung jedoch derer der Körperschaften.

Da in der Praxis Stiftungen mit gemeinnützigem Zweck deutlich überwiegen, soll dann intensiver auf die möglichen Steuerbegünstigungen von gemeinnützigen unternehmensverbundenen Stiftungen eingegangen werden. Hierzu muss eine Vielzahl an Voraussetzungen erfüllt werden, um als gemeinnützige Körperschaft anerkannt zu werden und in den Genuss der Vergünstigungen zu kommen. Neben der Mittelverwendung für gemeinnützige, mildtätige und kirchliche Zwecke müssen bei der Mittelverwendung gewisse Grundsätze eingehalten werden. Außerdem ergeben sich aus dem Gesetz formelle Voraussetzungen, wie z. B. Anforderungen an die tatsächliche Geschäftsführung der Stiftung. Bei der Besteuerung von gemeinnützigen Stiftungen wird zwischen den verschiedenen Einkommens- und Ausgabensphären unterschieden. So ergeben sich Unterschiede in der Besteuerung des ideellen Bereichs, des wirtschaftlichen Geschäftsbetriebs und der reinen Vermögensverwaltung. Eine besondere Rolle spielen dann auch die Zuwendungen an gemeinnützige Stiftungen, da diese im Rahmen des Sonderausgabenabzuges ggf. angerechnet werden können.

Zuletzt soll kurz auf eine besondere Gestaltungsform der Stiftung hingewiesen werden. Die Kombination einer Familienstiftung und einer gemeinnützigen Stiftung zur Errichtung eines Doppelstiftungsmodells. Hier werden die steuerlichen Vorteile der gemeinnützigen Stiftung mit den Wünschen der Stifterfamilie vereint.

2. Grundlagen

Für den Begriff der Stiftung liegt keine allgemeingültige Definition vor. Vielmehr lassen sich eine Reihe Gemeinsamkeiten der unterschiedlichen Stiftungsformen herausstellen, die als funktionelle Beschreibung dienen können: Eine Stiftung verfolgt einen fest definierten Zweck, welcher bei Gründung angegeben wird und der später nur noch schwer verändert werden kann. Die Stiftung ist mit ihrem angegebenen Zweck auf Dauer angelegt, in der Regel über den Tod des Stifters hinaus. Die Stiftung sollte deshalb mit einem ausreichenden Vermögen ausgestattet sein, damit die Erträge der Vermögensmasse für die Erfüllung des Stiftungszwecks ausreichen. Ein weiterer Aspekt ist die eigenständige Organisati-

on.³ Diese konstituierenden Merkmale werden als allgemein gültig angesehen und sind zudem gem. §80 Abs. 1 Satz 3 BGB neben Angabe zu Namen und Sitz der Stiftung notwendige Bestandteile der Stiftungssatzung.

Grundlagen für die gesetzlichen Regelungen zu Stiftungen finden sich in §§80 bis 88 BGB. Des Weiteren sind landesspezifische Gesetze vorhanden – im weiteren Verlauf wird auf das StiftG NRW verwiesen. Für steuerliche Fragen ist zum einen die AO (insb. §§14, 51 ff.) zurate zu ziehen, zum anderen finden die Bestimmungen des Körperschaftsteuergesetzes, des Gewerbesteuergesetzes, des Einkommensteuergesetzes und des Erbschaft- und Schenkungsteuergesetzes Anwendung.⁴

Unternehmensverbundene Stiftungen, auch Unternehmensstiftungen genannt, sind Oberbegriffe für Stiftungen im unternehmerischen Bereich. Zu unterscheiden sind hier die beiden Untergruppen der Unternehmensträgerstiftungen, die selbst ein Wirtschaftsunternehmen betreiben – in der Praxis jedoch eher selten vorkommen – und die Beteiligungsträgerstiftungen in Form der Stiftung & Co. KG sowie die Beteiligung an einer Kapitalgesellschaft.⁵ Unternehmensverbundene Stiftungen zählen zu den rechtsfähigen Stiftungen des Privatrechts bzw. zu den Stiftungen des bürgerlichen Rechts.

Im Falle der Unternehmensträgerstiftung wird aus Stiftung und Unternehmen eine Einheit gebildet. Das bedeutet, dass die Stiftung Trägerin bzw. Eigentümerin des Unternehmens ist und damit nach außen tätig wird. Das Unternehmen hat in diesem Moment keine eigene Rechtspersönlichkeit, da diese an die Stiftung übergeht. Die Stiftung ist Rechts- und Steuersubjekt. Sie gehört nur sich selbst, denn Anteilseigner, Gesellschafter oder Mitglieder sind nicht vorhanden.⁶

Eine Stiftung und Co. KG ist eine Personengesellschaft, bei der i. d. R. die Stiftung als Komplementärin beteiligt ist. Damit kann die Stiftung die anhaltende Erfüllung des Stiftungszwecks garantieren. Es besteht die Möglichkeit, eine zweite Stiftung als Kommanditistin einzusetzen. Steuerlich ergibt sich durch die Rechtsform der Stiftung und Co. KG ein Vorteil gegenüber der Kapitalgesellschaft, da Verluste aus der Beteiligung an dieser mit anderen positiven Einkünften verrechnet werden können.⁷

Bei einer Beteiligungsträgerstiftung handelt es sich um eine Stiftung, die selbst nicht gewerblich tätig wird, aber als Gesellschafterin bzw. Eigentümerin an einer Personen- oder Kapitalgesellschaft beteiligt ist.⁸ Das Unternehmen unterliegt den Vorschriften der jeweiligen Unternehmensrechtsform und kann daher, verglichen mit der Unternehmensfüh-

³Vgl. Meyn, Richter in Meyn et al. (2013), Rz. 2 so auch Schlüter and Stolte (2016), Kap. 1 Rz. 3 und Stumpf in Stumpf et al. (2011), Teil B Kap. 1 § Rz. 7.

⁴Vgl. Wigand in Wigand et al. (2015), Kap. 2 Rz. 6.

⁵Vgl. Richter, Gollan in Meyn et al. (2013), Rz. 105, so auch G. and T. (2014), S. 93.

⁶Vgl. G. and T. (2014), S. 93.

⁷Vgl. G. and T. (2014), S. 102 f.

⁸Vgl. G. and T. (2014), S. 117.

rung, durch die Stiftung flexibler bleiben.⁹ Es kann bspw. zu Problemen mit Arbeitnehmern des Unternehmens kommen, wenn bei einer Unternehmensträgerstiftung die Gewinne an die Stiftung bzw. später an ihre Destinatäre, die Begünstigten der Stiftung, übergehen. Besser zu vermitteln ist ggf. die Gewinnausschüttung an kapitalgebende Gesellschafter bzw. Aktionäre, wie es bei der Beteiligungsträgerstiftung der Fall ist.¹⁰ Eine enge Verbindung, wie bei der Unternehmensträgerstiftung, lässt sich auch bei der Beteiligungsträgerstiftung realisieren. Dies kann erreicht werden, indem Geschäftsführung des Unternehmens und Vorstand der Stiftung von derselben Person bekleidet werden.¹¹

2.1. Beweggründe und Motivation der Stifter

Hinter der Entscheidung eine Stiftung zu gründen können vielfältige Motive des Stifters liegen. Dabei wird jedoch zwischen dem Motiv eine unternehmensverbundene Stiftung zu gründen und dem in der Stiftungssatzung festgelegten Stiftungszweck unterschieden. Das Motiv des Stifters kann vereinfacht als Beweggrund für die Stiftung oder als Anliegen, das mit der Stiftung verbunden wird, definiert werden.¹²

Der wohl wichtigste Grund zur Errichtung einer unternehmensverbundenen Stiftung ist die Festlegung von Unternehmensnachfolgeregelungen. Der Stifter – also Gründer des Unternehmens – möchte die Unternehmenskontinuität gesichert wissen und handelt aus betriebswirtschaftlichem Interesse. Sein Unternehmen soll vor Einflüssen des Staates, fremden Interessensgruppen oder anderen Einzelpersonen, wie Familienmitgliedern, die versuchen andere Ziele als der Stifter zu verfolgen, geschützt werden.¹³

Einem möglichen Erbenstreit nach dem Tod des Unternehmers kann durch entsprechende Stiftungsgestaltung entgegengewirkt werden. Wird die Stiftung spätestens zehn Jahre vor dem Tod des Stifters gegründet und das Vermögen auf diese übertragen, bestehen im Erbfall keine Ansprüche, die zur Liquidation oder Teilung des Vermögens führen, da sich die Pflichtteilsansprüche nur auf den Nachlass des Erblassers außerhalb der Stiftung beziehen. Damit können die Erben bei der Unternehmensnachfolge das Unternehmen bzw. die Gesellschaftsanteile nicht einzeln veräußern.¹⁴ Sind noch keine zehn Jahre zwischen der Vermögensübertragung und dem Tod des Stifters vergangen, können die berechtigten Erben ihren Pflichtteil anteilig geltend machen. Es gilt das sog. „Abschmelzungsmodell“, bei dem die Pflichtteilsrelevanz der Stiftungsdotation jährlich ab dem Tag der Zuwendung um 10 Prozent abnimmt. So können nach einem Jahr nur noch 90 Prozent der Stiftungsdotation Pflichtteilsergänzungsansprüche auslösen.¹⁵

Ein weiterer Aspekt der Sicherung der Unternehmenskontinuität ist das Fehlen von geeigneten Erben für die Unternehmensnachfolge. Sind keine entsprechend qualifizierten Unternehmerpersönlichkeiten in der Folgegeneration oder fehlt die Bereitschaft zur Fortführung des Unternehmens, kann das Nachfolgeproblem durch eine Stiftung gelöst werden. Der Stiftungsgründer kann dadurch sicherstellen, dass ein geeigneter Nachfolger die Unternehmensführung übernimmt, in dem er in der Stiftungssatzung die Bestellung des Nachfolgers von Bedingungen abhängig macht.¹⁶

Neben dem wohl wichtigsten Motiv der Unternehmensfortführung können auch weitere eher ideelle Motivationslagen hinzukommen wie z. B. die Wahrung der Traditionen des Unternehmens bei einer Fusion oder dem Verkauf des Unternehmens. Es besteht auch die Möglichkeit, die bisherigen ehrenamtlichen Engagements aus Teilbereichen des Unternehmens in einer entsprechenden Stiftung zu bündeln.¹⁷

Auch andere ideelle Gründe, die sich häufig in gemeinnützigen Stiftungszwecken widerspiegeln, können ausschlaggebend zur Gründung einer Stiftung sein.¹⁸ Der Stifter möchte dann bspw. ein gesellschaftliches Anliegen, mit dem er sich identifiziert, fördern oder aber dieser will seinem Verantwortungsbewusstsein gegenüber der Gesellschaft gerecht werden und dieser etwas zurückgeben, nachdem es ihm selbst gut ergangen ist.¹⁹

Monetäre Gründe, wie z. B. die Reduzierung der Steuerlast durch steuerliche Optimierungsmöglichkeiten, können eine weitere Rolle spielen, wobei ihnen aber eine geringere Bedeutung zuzuweisen ist als den zuvor erläuterten Motiven.²⁰

2.2. Entstehung einer rechtsfähigen Stiftung

Damit eine Stiftung die Rechtsfähigkeit erlangen kann, müssen zwei Voraussetzungen erfüllt werden. Gem. §80 Abs. 1 BGB ist zum einen das Stiftungsgeschäft erforderlich, zum anderen muss die Stiftung durch die zuständige Landesbehörde anerkannt werden. Die Voraussetzungen stehen selbstständig nebeneinander.²¹

Beim Stiftungsgeschäft handelt es sich um eine einseitige, nicht empfangsbedürftige Willenserklärung. Es wird innerhalb des Stiftungsgeschäfts zwischen einem organisatorischen und einem vermögensrechtlichen Teil unterschieden. Der organisatorische Teil bestimmt gem. §85 BGB die Verfassung der Stiftung, soweit sie nicht auf Bundes- oder Landesgesetz beruht. Der vermögensrechtliche Teil bezieht sich auf die Zuwendung von Vermögen zum Stiftungsvermögen zur Verfolgung des Stiftungszwecks.²² Diese beiden Teile werden inhaltlich in der Stiftungsabsicht des Stifters festgehalten. Formerfordernisse und notwendige Inhalte werden für

⁹Vgl. Saenger in Werner and Saenger (2008), Rz. 195.

¹⁰Vgl. G. and T. (2014), S. 118.

¹¹Vgl. G. and T. (2014), S. 117.

¹²Vgl. Engel (2008), S. 92.

¹³Vgl. Engel (2008), S. 93 f.

¹⁴Vgl. Engel (2008), S. 94 f.

¹⁵Vgl. Wigand in Wigand et al. (2015) Kap. 10 Rz. 25

¹⁶Vgl. Engel (2008), S. 97.

¹⁷Vgl. Haase-Theobald in Wigand et al. (2015) Kap. 1 Rz. 36

¹⁸Vgl. Fleschutz (2008), S. 104 i. V. m. Schmitzer (2006), S. 18.

¹⁹Vgl. Haase-Theobald in Wigand et al. (2015) Kap. 1 Rz. 35

²⁰Vgl. Götz in Götz and PachHanssenheimb (2016), Rz. 25 i. V. m. Fleschutz (2008), S. 104.

²¹Vgl. Stumpf in Stumpf et al. (2011), Teil B Kap. 2 § 80 BGB Rz. 4.

²²Vgl. Stumpf in Stumpf et al. (2011), Teil B Kap. 2 § 80 BGB Rz. 5.

Stiftungsgeschäfte unter Lebenden in §81 BGB und für Stiftungsgeschäfte von Todes wegen in §83 BGB i. V. m. den erbrechtlichen Bestimmungen geregelt.²³ Teil des Stiftungsgeschäfts ist auch die angesprochene Stiftungssatzung, die gem. §81 Abs. 1 BGB Angaben zum Namen, zum Sitz, zum Zweck, zum Vermögen und zur Bildung des Vorstandes der Stiftung enthalten muss.

Liegt neben dem Stiftungsgeschäft auch die Anerkennung durch die zuständige Behörde des Landes vor, entsteht die rechtsfähige Stiftung. Die Anerkennung als solche stellt einen privatrechtsgestaltenden Verwaltungsakt dar. Voraussetzungen für die Anerkennung durch die Behörden ist zunächst das angesprochene Stiftungsgeschäft, außerdem die Verfolgung eines zulässigen Stiftungszwecks mit einem ausreichenden Stiftungsvermögen und zuletzt die Konformität des Stiftungszwecks mit dem Gemeinwohl.²⁴

Lange Zeit war umstritten, ob unternehmensverbundene Stiftungen anerkannt werden können. Vor allem die Frage, welchen Zweck solche Stiftungen verfolgen, war nicht geklärt. Seit der Einführung des Gesetzes zur Modernisierung des Stiftungsrechts sind Unternehmensstiftungen grundsätzlich rechtlich zulässig und deshalb durch die zuständige Landesbehörde anzuerkennen.²⁵ Zunächst gab es Zweifel daran eine unternehmensverbundene Stiftung bestätigen zu können, da der übereinstimmende Zweck des Unternehmens und der Stiftung eine Problematik darstellt.²⁶ Besteht der alleinige Zweck einer Stiftung darin ein Unternehmen zu führen, handelt es sich um eine sog. Selbstzweckstiftung, die nicht zulässig ist.²⁷

Natürlich müssen auch unternehmensverbundene Stiftungen die allgemeinen Voraussetzungen zur Anerkennung erfüllen. Da Unternehmensstiftungen keine eigene Rechtsform bilden, sondern lediglich die Anwendung der Stiftungsrechtsform auf ein Unternehmen darstellen, gelten für sie die allgemeinen Regelungen des Stiftungsrechts.

Vor allem bei unternehmensverbundenen Stiftungen ist der Zeitpunkt der Errichtung der Stiftung sehr relevant. Fällt die Errichtung der Stiftung auf den Todeszeitpunkt des Unternehmers oder ist diese erst kurz zuvor geplant, kann es zu Problemen in der Unternehmensnachfolge kommen. Die grundsätzliche Problematik eines möglichen Bruchs beim Tod des Unternehmers kann ggf. verstärkt werden, denn das Aneignen der notwendigen Fähigkeiten, um Unternehmer- und Eigentümerfunktionen der Stiftung zu übernehmen, braucht Zeit.²⁸

²³Vgl. Stumpf in Stumpf et al. (2011), Teil B Kap. 2 § 80 BGB Rz. 23.

²⁴Vgl. Stumpf in Stumpf et al. (2011), Teil B Kap. 2 § 80 BGB Rz. 38 ff. 65.

²⁵Vgl. Pöllath, Richter in Seifart (2009), § 12 Rz. 136, so auch Saenger in Werner and Saenger (2008), Rz. 876.

²⁶Vgl. Pöllath, Richter in Seifart (2009), § 12 Rz. 53.

²⁷Vgl. Pöllath, Richter in Seifart (2009), § 12 Rz. 142.

²⁸Vgl. Richter, Gollan in Meyn et al. (2013), Rz. 106.

2.2.1. Stiftungszweck

Der Stiftungszweck ist der zentrale Bestandteil einer jeden Stiftung, daher wird er auch als die „Seele der Stiftung“²⁹ bezeichnet. Erst durch ihn erhält das Vermögen eine rechtliche Existenz in Form der Stiftung.³⁰ Da die Organe einer Stiftung ausschließlich zur Erfüllung des Zwecks dienen und nicht ihren eigenen Willen verfolgen dürfen, ergeben sich aus dem Stiftungszweck Leitlinien für ihre Stiftungstätigkeit. Der Stiftungszweck gibt den Organen einen klar abgegrenzten und eindeutigen Auftrag und verhindert jegliche Willkür der Organe.

Ist die Stiftung mit ihrem Zweck einmal anerkannt, finden Änderungen des Stiftungszwecks nur unter besonderen Voraussetzungen statt. Gem. §87 Abs. 1 BGB kann der Stiftungszweck durch die zuständige Behörde angepasst werden, falls dieser nicht mehr erfüllt werden kann oder er das Gemeinwohl gefährdet.

Grundsätzlich kann eine Stiftung alle Arten von Zwecken verfolgen – meist werden altruistische Zwecke gewählt, aber auch wirtschaftliche oder privatnützige Zwecke stehen einer Anerkennung nicht entgegen.³¹ Einzige Voraussetzungen sind, dass der Zweck nicht dem Gemeinwohl entgegensteht, also gegen geltendes Gesetz oder die guten Sitten verstößt, oder die Erfüllung des Zwecks dauerhaft unmöglich ist.³² Als Stiftungszweck ausgeschlossen ist zudem der reine Selbstzweck. Die ausschließliche Erhaltung und Bewirtschaftung des eigenen Vermögens, die alleinige Erfüllung der Interessen des Stifters bzw. dass der Stifter selbst Nutznießer der von ihm errichteten Stiftung ist, ist somit nicht zulässig.³³

Bei der Auswahl des Stiftungszwecks muss sich der Stifter nicht auf einen einzelnen Zweck beschränken. Grundsätzlich ist es möglich auch mehrere Stiftungszwecke zu benennen, die entweder durch den Stifter selbst oder durch den Vorstand der Stiftung unterschiedlich priorisiert werden können.³⁴

Das mit der Stiftung verbundene Unternehmen kann zur Mittelerzielung der Stiftung dienen oder aber als Zweck des Mitteleinsatzes der Stiftung gesehen werden.

Eine Unternehmensträgerstiftung verfolgt meist sowohl einen wirtschaftlichen Zweck – nämlich das Fortbestehen des Unternehmens durch Erzielung von Unternehmenserträgen – als auch einen ideellen Zweck. Dieser kann sehr unterschiedlich ausfallen. In Frage kommen neben innerbetrieblichen Zwecken wie die Förderung der Mitarbeiter auch außerbetriebliche Zwecke wie die Förderung der Forschung und Entwicklung in dem Bereich, in dem das Unternehmen tätig ist.³⁵ Bei Unternehmensträgerstiftungen ist es sinnvoll eine Klausel in die Satzung mit aufzunehmen, die es dem Vorstand bzw. anderen Gremien erlaubt den Stiftungszweck anzupassen.

²⁹Nissel in Werner and Saenger (2008), Rz. 207.

³⁰Vgl. Schlüter and Stolte (2016), Kap. 2 Rz. 47.

³¹Vgl. Götz in Götz and PachHanssenheimb (2016), Rz. 181.

³²Vgl. Schlüter and Stolte (2016), Kap. 2 Rz. 47.

³³Vgl. Götz in Götz and PachHanssenheimb (2016), Rz. 182-183.

³⁴Vgl. Stumpf in Stumpf et al. (2011), Teil B Kap. 2 § 80 BGB Rz. 47.

³⁵Vgl. G. and T. (2014), S. 94 f.

sen, wenn dies nötig ist, um das Unternehmen den marktwirtschaftlichen Erfordernissen anzupassen. Damit soll gewährleistet sein, dass die Leistungsfähigkeit des Unternehmens erhalten bleibt.³⁶

Die Stiftung und Co. KG wird als Stiftungszweck regelmäßig die Übernahme der geschäftsführenden Tätigkeiten des Personenunternehmens im Interesse der Unternehmerfamilie verfolgen und damit die Erhaltung des Unternehmens gewährleisten.³⁷ Wird als weiterer Stiftungszweck das Halten der KG-Anteile festgelegt, kann erreicht werden, dass die Stiftung und Co. KG praktisch übernahmefest ist. Dazu muss in der Satzung zusätzlich festgehalten werden, dass für Änderungen des Vertrages der KG die Zustimmung der Stiftung notwendig ist. Ein neutral besetzter Stiftungsrat kann eingesetzt werden, um einen möglichen Substanzentzug aus der KG durch Familienmitglieder zu verhindern.³⁸

Da die Beteiligungsträgerstiftung mit dem Unternehmen weniger eng verbunden ist als die Unternehmensträgerstiftung wird sie als Stiftungszweck nicht den wirtschaftlichen Aspekt der Erzielung von Unternehmenserträgen verfolgen. Bei der Ausgestaltung einer Familienstiftung wird die Versorgung der Familie des Stifters bzw. Unternehmers im Vordergrund stehen, oder es werden ideelle oder gemeinnützige Zwecke von der Beteiligungsträgerstiftung verfolgt.

2.2.2. Stiftungsvermögen

Neben dem Stiftungszweck spielt auch das Stiftungsvermögen eine entscheidende Rolle, da dies der Zweckerfüllung dient. Gem. §80 Abs. 2 BGB muss für die Anerkennung einer Stiftung vorausgesetzt sein, dass die Erträge des gestifteten Vermögens auch ausreichen, um den Zweck dauerhaft zu verfolgen. Weitere Angaben zu Art und Höhe des notwendigen Vermögens macht der Gesetzgeber jedoch nicht.

Grundsätzlich kann ein Stifter sowohl Bar- als auch Sachwerte in seine Stiftung einbringen. Es ist jedoch zu berücksichtigen, dass Stiftungen, die lediglich Sachwerte zu ihrem Stiftungsvermögen zählen, selten überlebensfähig sind. Hier besteht meist die Notwendigkeit weiterer Zuwendungen in Form liquider Mittel zur Existenzsicherung der Stiftung, denn laufende Aufwendungen, wie z. B. Versicherungsaufwendungen der Sachwerte, müssen gedeckt werden.³⁹ Sowohl Bar- und Wertpapiere als auch Immobilien, bewegliche Sachen sowie Rechte, wie bspw. Patente, Forderungen und Unternehmensbeteiligungen, können als Vermögenswerte in die Stiftung eingebracht werden.

Eine Mindesthöhe des Stiftungsvermögens ist zwar gesetzlich nicht festgelegt, in der Praxis wird jedoch zur Anerkennung der Stiftung meist ein Vermögen von mindestens 50.000 € angesetzt, damit der Stiftungszweck angemessen verfolgt werden kann. Hierzu ist es angebracht, die Relation von Zweck und Mittel kritisch zu hinterfragen.⁴⁰ Das Stiftungsvermögen besteht zum einen aus dem Anfangs- bzw.

Ausstattungsvermögen, das die Stiftung bei Gründung erhält und aus weiteren Zuwendungen, auch Zustiftungen genannt, die der Stiftung zur Erhöhung des Stiftungsvermögens zur Verfügung gestellt werden. Die Stiftungsmittel dagegen sind diejenigen Erträge, die aus der Verwaltung des Stiftungsvermögens an die Stiftung zurückfließen, wie bspw. Zinsen, Dividenden, Miet- und Pächterträge. Gem. §55 Abs. 1 AO sind diese zeitnah zu verbrauchen. Auch Zuwendungen, die nicht das Stiftungsvermögen erhöhen sollen, wie z. B. Spenden, zählen hierzu.⁴¹

Das Stiftungsvermögen der Unternehmensträgerstiftungen besteht aus dem Unternehmenskapital und umgekehrt. Für steuerliche Fragen wird zwischen Betriebs- und Privatvermögen unterschieden. Bspw. zählen nicht betrieblich genutzte Grundstücke zum Privatvermögen und eine Veräußerung wäre ertragsteuerfrei.⁴²

Bei Beteiligungsträgerstiftungen besteht das Stiftungsvermögen meist ausschließlich oder in wesentlichen Teilen aus der Beteiligung an dem verbundenen Unternehmen, dessen Gewinne der Stiftung zufließen.⁴³

Neben der ersten Vermögensausstattung der Stiftung bei Gründung spielt auch die Kapitalbeschaffung eine große Rolle im Hinblick auf die unternehmensverbundenen Stiftungen. Unternehmensträgerstiftungen sind bei der Kapitalbeschaffung, um je nach Markt- und Unternehmenslage Investitionen vornehmen zu können, eher unflexibel. Durch die fehlende Fähigkeit, Kapitalerhöhungen durchzuführen, wird die Handlungsfähigkeit der Stiftung eingeschränkt. Beteiligungsträgerstiftungen sind hier besser aufgestellt, denn die Führung des Unternehmens in eigener Rechtsform ermöglicht es dem Unternehmen, Rücklagen zu bilden oder Fremdmittel aufzunehmen.⁴⁴

2.2.3. Stiftungsorganisation

Damit die Stiftung die Handlungsfähigkeit erlangt, muss eine entsprechende Stiftungsorganisation errichtet werden. Die Organisation ist in der Stiftungssatzung, der Stiftungsverfassung oder auch dem Stiftungsstatut verankert. Die Satzung enthält neben Angaben zu den unterschiedlichen Gremien und Organen der Stiftung auch Aussagen zu Geschäftsbereichen, Vertretungsmacht, Einberufung, Beschlussfähigkeit ihrer Gremien und zu Vergütungen der Vertreter.⁴⁵ Die Organisation einer Stiftung ist verantwortlich für die Umsetzung des Stiftungszwecks.

Die gesetzliche Mindestanforderung an die Organisation einer Stiftung liegt nur in der Konstitution eines Vorstandes. Dieser ist gem. §86 i. V. m. §26 Abs. 2 Satz 1 BGB der gesetzliche Vertreter und vertritt die Stiftung gerichtlich und außergerichtlich. Es bedarf also mindestens einer natürlichen oder juristischen Person, die die Stiftung als Vorstand führt und vertritt.

³⁶Vgl. G. and T. (2014), S. 98.

³⁷Vgl. G. and T. (2014), S. 111.

³⁸Vgl. G. and T. (2014), S. 103.

³⁹Vgl. Haase-Theobald in Wigand et al. (2015), Kap. 7 Rz. 3.

⁴⁰Vgl. Schlüter and Stolte (2016), Kap. 5 Rz. 1.

⁴¹Vgl. Schlüter and Stolte (2016), Kap. 5 Rz. 2.

⁴²Vgl. G. and T. (2014), S. 94.

⁴³Vgl. Engel (2008), S. 41.

⁴⁴Vgl. Saenger in Werner and Saenger (2008), Rz. 898.

⁴⁵Vgl. Berndt in Berndt and Götz (2009), Rz. 311.

Neben dem Vorstand können weitere Organe durch die Stiftungsverfassung vorgesehen werden. Gerade bei größeren Stiftungen spielen Stiftungsrat, Kuratorium, Verwaltungsrat o. Ä. eine wichtige Rolle, um als Kontrollgremium zu agieren.⁴⁶ Zusätzlich zur Funktion der Überwachung kann der Stiftungsrat durch die Stiftungssatzung dazu befähigt werden, den Vorstand zu bestellen und abzurufen, wie es auch beim aktienrechtlichen Aufsichtsrat gem. §84 AktG der Fall ist. Die Bestellung eines Stiftungsrates kann der Einbindung von fachlichem Sachverstand, persönlicher Kompetenz und äußerem Ansehen dienen.⁴⁷

Vor allem bei unternehmensverbundenen Stiftungen spielt die organisatorische Gestaltung der Stiftung eine besondere Rolle: Handelt es sich um eine Unternehmensträgerstiftung, bei der die Stiftung selbst ein Wirtschaftsunternehmen betreibt, gilt es besondere Anforderungen an die Stiftungsverfassung zu stellen. Die übliche Organisation einer Stiftung ist zur Führung eines Unternehmens strukturell eher ungeeignet. Insbesondere die Vertretung der Stiftung durch den Vorstand ist problematisch, wenn dieser nicht über die unternehmerischen Qualitäten zur Führung des Unternehmens verfügt. In diesem Fall müssen in der Stiftungsverfassung Regelungen verankert werden, wie der Vorstand berufen und abgerufen wird.⁴⁸ Da die Stiftung nicht wie bei einer Gesellschaft zugehörige Gesellschafter hat, die üblicherweise Kontroll- und Eigentümerrechte ausüben, ist es sinnvoll getrennte Organe für Geschäftsführung und Kontrollfunktionen einzuführen. Die Trennung wird auch erreicht, indem das Unternehmen unterhalb der Stiftung organisiert ist und die Stiftung als Gesellschafter des Unternehmens agiert, wie es bei der Beteiligungsträgerstiftung der Fall ist.⁴⁹

2.3. Stiftungsaufsicht über rechtsfähige Stiftungen

Neben der notwendigen Anerkennung der Stiftung durch die zuständige Landesbehörde wirkt der Staat auch während des Bestehens einer Stiftung als Stiftungsaufsicht mit.⁵⁰ Begründet wird diese Beaufsichtigung durch die spezielle Organisationsstruktur der Stiftung. Mit der Anerkennung der Rechtsfähigkeit existiert die Stiftung unabhängig vom Stifter, der keinen Einfluss mehr auf die Handlung der Stiftung hat, es sei denn er bestimmt sich selbst als Vorstand. Zur Erfüllung des Stifterwillens ist die Stiftungsaufsicht deshalb verpflichtet und befugt, auf die Einhaltung stiftungsrechtlicher Bestimmungen hinzuwirken.⁵¹

Da rechtsfähige Stiftungen juristische Personen i. S. d. Art. 19 Abs. 3 GG sind, sind sie Träger der Grundrechte und genießen Grundrechtsschutz. Die Eingriffsmöglichkeiten der Stiftungsaufsicht sind somit insofern beschränkt, dass diese an die allgemeinen Prinzipien des Verwaltungsrechts gebunden ist. Dazu zählen bspw. das Subsidiaritätsprinzip und der

Verhältnismäßigkeitsgrundsatz. Nach dem Subsidiaritätsprinzip darf die Stiftungsbehörde nur insoweit eingreifen, wie die Stiftung selbst keine ausreichenden Maßnahmen zur Erfüllung der rechtlichen Vorgaben vorsieht. Der Verhältnismäßigkeitsgrundsatz verlangt von der Aufsichtsbehörde, nur solche Mittel einzusetzen, die unbedingt erforderlich und angemessen sind. Kann ein rechtmäßiges Ziel mit unterschiedlichen Maßnahmen erreicht werden, so ist das mildeste Mittel zu wählen.⁵²

Die Stiftungsgesetze der Länder bestimmen, wer mit den Aufgaben der staatlichen Stiftungsaufsicht betraut wird.⁵³ In NRW übernimmt jeweils die Bezirksregierung, in deren Gebiet sich der Sitz der Stiftung befindet, als sog. Stiftungsbehörde diese Funktion.⁵⁴ Die Stiftungsbehörde stellt dabei nach §6 Abs. 2 StiftG NRW sicher, dass der Wille des StifTERS, der aus Stiftungsgeschäft und Stiftungssatzung hervorgeht, beachtet wird und die Tätigkeit der Stiftung nicht gegen Recht und Gesetz verstößt. Gem. §6 Abs. 3 StiftG NRW unterliegen zwar auch privatnützige Stiftungen der Stiftungsaufsicht, doch in weit geringerem Umfang.⁵⁵ Die Stiftung wird teilweise aus der Aufsicht entlassen, denn die Stiftungsbehörde prüft lediglich, dass die Tätigkeit der Stiftung nicht gegen die gesetzlich geschützten öffentlichen Interessen verstößt.⁵⁶

Die Stiftungsaufsicht agiert als reine Rechtsaufsicht. Das bedeutet, dass sie nicht an Stelle der Stiftungsorgane Entscheidungen treffen darf. Die Aufsicht behandelt weiterhin lediglich stiftungsspezifische Fragen. Das heißt, dass bei Prüfung der Einhaltung der Rechtsordnung nur Stiftungsverfassung, die §§80 ff. BGB und das entsprechende Landesstiftungsrecht gemeint sind – nicht jedoch privatrechtliche Vorschriften.⁵⁷

Neben der Kontrollfunktion im Hinblick auf die Erfüllung des Stifterwillens, kommt die Stiftungsaufsicht häufig auch beratenden Funktionen nach. Diese sind zwar nicht im StiftG NRW gesetzlich festgehalten, in der Regel erhalten angehende Stifter und ihre Berater jedoch Empfehlungen und Hinweise zu kritischen Fragen bspw. während der Stiftungserichtung, der Satzungsgestaltung oder der notwendigen Vermögensausstattung.⁵⁸ Damit soll bei Maßnahmen, bei denen die Behörde zustimmen muss, das Verfahren beschleunigt werden bzw. soll Verstößen gegen die Rechtsordnung vorgebeugt werden.⁵⁹

Zur Erfüllung ihrer Funktionen als Stiftungsaufsicht kann die Behörde verschiedene vorbeugende und repressive Maßnahmen durchführen. Zu den präventiven Maßnahmen zählen bspw. Maßnahmen der Überwachung, Beratung und Unterrichtung wie z. B. den behördlichen Genehmigungsvorbehalten. Repressive Maßnahmen kommen dann in Betracht,

⁵²Vgl. Schlüter and Stolte (2016), Kap. 3 Rz. 3.

⁵³Vgl. Stolte in Wigand et al. (2015), Kap. 3 Rz. 104.

⁵⁴Vgl. Ministerium für Inneres und Kommunales des Landes Nordrhein-Westfalen, Stiftungsaufsicht.

⁵⁵Vgl. OFD NRW (2017), S. 20.

⁵⁶Vgl. Stolte in Wigand et al. (2015), Kap. 3 Rz. 107.

⁵⁷Vgl. Schlüter and Stolte (2016), Kap. 3 Rz. 4.

⁵⁸Vgl. Stolte in Wigand et al. (2015), Kap. 3 Praxishinweis nach Rz. 109.

⁵⁹Vgl. Schlüter and Stolte (2016), Kap. 3 Rz. 5.

⁴⁶Vgl. Berndt in Berndt and Götz (2009), Rz. 312.

⁴⁷Vgl. Götz in Götz and PachHanssenheimb (2016), Rz. 238.

⁴⁸Vgl. Götz in Götz and PachHanssenheimb (2016), Rz. 29 Nr. 3.

⁴⁹Vgl. Pöllatz, Richter in Seifart (2009), § 12 Rz. 174 f.

⁵⁰Vgl. OFD NRW (2017), S. 20.

⁵¹Vgl. Schlüter and Stolte (2016), Kap. 3 Rz. 2.

wenn die Stiftung gegen ein Gesetz oder ihre Satzung verstößt. Die Stiftungsaufsicht kann dann durch Beanstandungen, Aufhebung von Maßnahmen und Anordnungen eingreifen.⁶⁰

Verletzt die Stiftungsaufsicht ihre Amtspflicht i. S. d. Art. 34 GG und erleidet die Stiftung dadurch einen Schaden, so kann sie das Bundesland als Träger der Stiftungsaufsicht gem. §839 BGB auf Schadensersatz aus Amtshaftung verklagen.⁶¹ Hierzu kann es bspw. kommen, wenn die Behörde nicht einschreitet, obwohl sie von einer zweckwidrigen Verwendung der Stiftungsmittel Kenntnis hatte. Ein Amtshaftungsanspruch kann auch entstehen, wenn ein Organmitglied Gelder der Stiftung veruntreut, denn zu den Pflichten der Aufsicht zählt auch die sorgsame Prüfung des Jahresabschlusses und anderer Finanzberichte.⁶² Trägt der Stiftungsvorstand eine Mitschuld bei der Entstehung eines Schadens, ist §254 BGB anzuwenden und der Vorstand kann für diese belangt und zu Schadenersatzzahlungen verpflichtet werden.⁶³

3. Besteuerung unternehmensverbundener Stiftungen und ihrer Begünstigten

3.1. Besteuerung bei Stiftungserrichtung

Die Besteuerung von unternehmensverbundenen Stiftungen, die privatnützige Zwecke verfolgen, ergibt sich zum einen aus der Steuerpflicht der Stiftung als Steuersubjekt, zum anderen aus der steuerpflichtigen Übertragung von Vermögenswerten zwischen Steuersubjekten, also zwischen Stifter und Stiftung. Sie ähnelt sehr der Besteuerung anderer Körperschaften.

Das zur Anerkennung notwendige Stiftungsvermögen wird bei Stiftungserrichtung vom Stifter an die Stiftung übertragen. Die dabei übertragenen Werte scheiden beim Errichter der Stiftung aus und sind dann der Stiftung zuzurechnen.⁶⁴ In sachlicher Hinsicht führt die Errichtung der Stiftung also zu einer unentgeltlichen Zuwendung von Vermögenswerten.⁶⁵ Diese Zuwendung ist erbschaftsteuerpflichtig gem. §1 Abs. 1 ErbStG i. V. m. §3 Abs. 2 Nr. 1 ErbStG durch den Erwerb von Todes wegen, wenn die Stiftung von Todes wegen errichtet wird, oder schenkungsteuerpflichtig gem. §7 Abs. 1 Nr. 8 Satz 1 ErbStG durch die Schenkung unter Lebenden, wenn der Stifter bereits zu Lebzeiten die Stiftung errichtet und mit Vermögen ausstattet.⁶⁶ Gem. §2 Abs. 1 ErbStG entsteht die persönliche Steuerpflicht, wenn entweder der Stifter als Erblasser bzw. Schenker oder die Vermögen empfangende Stiftung ihren Wohnsitz gem. §8 AO, ihren gewöhnlichen Aufenthalt gem. §9 AO oder ihren Sitz gem. §11

AO im Inland hat.⁶⁷ Die persönliche Steuerpflicht gilt auch, wenn diese Voraussetzungen zwar nicht erfüllt sind, es sich aber um Inlandsvermögen i. S. d. §121 BewG handelt.⁶⁸

Der Zeitpunkt der Steuerpflicht liegt grundsätzlich im Übertragungsakt.⁶⁹ Nach §9 Abs. 1 Nr. 1 und 2 ErbStG entsteht die Steuerpflicht bei einem Erwerb von Todes wegen mit dem Tod des Erblassers, bei Schenkungen unter Lebenden mit Ausführung der Zuwendung. Diejenigen Vermögensgegenstände, die nach dem Tod des Stifters, aber vor Anerkennung der Stiftung durch die Stiftungsaufsicht, also auch vor Rechtsfähigkeit der Stiftung an diese übertragen werden, sind nach §1 ErbStG i. V. m. §9 Abs. 1 Nr. 1c ErbStG der steuerbaren Erstaussstattung zuzurechnen, da mit Anerkennung der Stiftung die Erbschaftsteuerpflicht entsteht.⁷⁰ Gleiches gilt für die Erträge, die in diesem Zeitraum erwirtschaftet wurden. Sie sollen in den Gesamtwert des steuerbaren Erwerbs eingerechnet werden.⁷¹ Allerdings kann dieses Vorgehen zu einer Doppelbelastung desselben Einkommens mit Ertragsteuer und Erbschaftsteuer führen und ist deshalb nicht unumstritten.⁷²

Besteuert wird nach §10 Abs. 1 ErbStG grundsätzlich der Wert der Bereicherung am Stichtag der Steuerpflicht, soweit diese nicht durch weitere gesetzliche Regelungen steuerfrei ist. Steuerbefreiungen können sich aus der Art des zugewendeten Vermögens ergeben (z. B. Kunst- oder wissenschaftliche Sammlungen gem. §13 Abs. 1 ErbStG) oder aufgrund von Freibeträgen nach §16 ErbStG. Die Stiftung wird bei Errichtung und Vermögensübertragung um den Nettowert ihrer Zuwendungen bereichert, also dem Mehrwert der Aktiva über die Passiva.⁷³ Der Wert dieser steuerpflichtigen Bereicherung wird gem. §12 Abs. 1 ErbStG nach den Vorschriften des Bewertungsgesetzes ermittelt. Die Bemessungsgrundlage für die Erbschaftsteuer nach §10 Abs. 5 ErbStG wird um Nachlassverbindlichkeiten gemindert. Dazu zählen Erblasserschulden, Erbfallschulden und sonstige Nachlassverbindlichkeiten i. S. d. §10 Abs. 5. Nr. 1 bis 3 ErbStG.

Zählt zu dem übertragenen Vermögen Betriebsvermögen eines Unternehmens, so sind die gesetzlichen Regelungen aus §13a und §13b ErbStG zu berücksichtigen. Es ergeben sich steuerliche Begünstigungen aus §13a ErbStG. Dabei zählt zum steuerlich begünstigungsfähigen Vermögen land- und forstwirtschaftliches Vermögen nach §13b Abs. 1 Nr. 1 ErbStG, inländisches Betriebsvermögen laut §13b Abs. 1 Nr. 2 ErbStG und Anteile an Kapitalgesellschaften gem. §13b Abs. 1 Nr. 3 ErbStG, sofern der Stifter am Nennkapital dieser Gesellschaft mit mehr als 25 Prozent unmittelbar beteiligt ist. Das begünstigte Vermögen ergibt sich gem. §13b Abs. 2 ErbStG aus der Differenz des begünstigungsfähigen Vermögens und des steuerschädlichen Verwaltungsvermögens. Bei

⁶⁰Vgl. Schlüter and Stolte (2016), Kap. 3 Rz. 10, 19.

⁶¹Vgl. Schlüter and Stolte (2016), Kap. 3 Rz. 47.

⁶²Vgl. Meyn in Meyn et al. (2013), Rz. 649 f.

⁶³Vgl. BGH vom 03.03.1977, III ZR 10/74, BGHZ 68, S. 142.

⁶⁴Vgl. Pöllath, Richter in Seifart (2009), § 40 Rz. 1.

⁶⁵Vgl. Richter, Gollan in Meyn et al. (2013), Rz. 542.

⁶⁶Vgl. Pöllath, Richter in Seifart (2009), § 40 Rz. 8 und Richter, Gollan in Meyn et al. (2013), Rz. 543 f.

⁶⁷Vgl. Richter, Gollan in Meyn et al. (2013), Rz. 546.

⁶⁸Vgl. G. and T. (2014), S. 71.

⁶⁹Vgl. Pöllath, Richter in Seifart (2009), § 40 Rz. 11.

⁷⁰Vgl. Richter, Gollan in Meyn et al. (2013), Rz. 560.

⁷¹Vgl. BFH vom 25.10.1995, II R 20/92, BStBl II 1996, S. 99 ff.

⁷²Vgl. Gebel (2002), Rz. 1233.

⁷³Vgl. Pöllath, Richter in Seifart (2009), § 40 Rz. 25 f.

Erfüllung der Voraussetzungen der §§13a und 13b ErbStG werden gesetzliche Bewertungsabschläge, die sog. Verschonungsabschläge i. H. v. 85 bzw. 100 Prozent gewährt.

Zusätzlich zu den Begünstigungen nach §13a ErbStG können ggf. weitere Steuerbefreiungen geltend gemacht werden. Hier sind der Verschonungsabschlag bei Großerwerb nach §13c ErbStG, der Entlastungsbetrag nach §19a ErbStG, das Erlass-Modell nach §28a ErbStG und die Stundungsregelung nach §28 ErbStG zu nennen.

Gem. §15 Abs. 1 ErbStG fällt die Zuwendung an eine rechtsfähige Stiftung in Steuerklasse III und damit in die ungünstigste Steuerklasse. Zuwendungen in diese erhalten nach §16 Abs. 1 Nr. 7 ErbStG einen Freibetrag von 20.000 € , der von der Zuwendung abgezogen werden darf. Der Steuersatz beträgt gem. §19 ErbStG 30 Prozent bzw. 50 Prozent bei einem steuerpflichtigen Erwerb von mehr als 6 Mio. € . Die Stiftung als Steuerschuldner, bzw. bei Schenkungen auch Stifter oder Zustifter, hat die Steuer nach Bekanntgabe des Steuerbescheides zu entrichten. Die Erbschaftsteuer wird gewöhnlich vom Bedachten, also der Stiftung, übernommen.⁷⁴

Ein Sonderfall innerhalb der erbschaftsteuerlichen Beurteilung von Vermögensübertragungen an Stiftungen ist die Übertragung an eine Familienstiftung. Eine Familienstiftung ist gem. §1 Abs. 1 Nr. 4 ErbStG eine Stiftung, die wesentlich im Interesse einer Familie oder bestimmter Familien errichtet worden ist. Von wesentlichem Interesse kann ausgegangen werden, wenn mehr als die Hälfte der Leistungen dem Stifter bzw. seinen Angehörigen zufließen.⁷⁵ Die Bestimmung der Steuerklasse und die damit verbundenen steuerlichen Folgen der Erbschaftsteuer werden für Familienstiftungen gesondert geregelt. Eine Verwandtschaft zwischen der Stiftung als Körperschaft und dem Stifter ist ausgeschlossen, doch für die Ermittlung der Steuerklasse erfolgt der Durchgriff auf die begünstigten Familienmitglieder. Das bedeutet, dass der Familie das sog. Steuerklassenprivileg nach §15 Abs. 2 Satz 1 ErbStG zu Teil wird. Maßgeblich ist hierbei das Verwandtschaftsverhältnis des nach der Stiftungsurkunde entferntesten Berechtigten zum Stifter. Dadurch fällt, je nach Verwandtschaftsverhältnis zu den Destinären, die Besteuerung in Steuerklasse I oder II.⁷⁶

Für den Stifter wird die unentgeltliche Übertragung von Vermögenswerten zunächst wie eine Veräußerung behandelt. Es kann dabei entweder zu einer Steuerpflicht auf den realisierten Gewinn oder aber zum steuerlichen Abzug von Betriebsausgaben, Werbungskosten oder Sonderausgaben in Form einer Spende kommen.⁷⁷ Die ertragsteuerlichen Folgen für den Stifter sind also abhängig davon, ob er durch die Übertragung von Vermögensgegenständen im Zuge der Stiftungserrichtung einen Gewinn realisiert.⁷⁸

Überträgt der Stifter Teile seines Privatvermögens, so ist dies meist keine steuerpflichtige Veräußerung unabhängig

davon, ob die Übertragung zu Lebzeiten oder von Todes wegen erfolgt. Denn der Stifter wird das Vermögen regelmäßig ohne Gegenleistung auf die Stiftung übertragen, sodass keine Gewinnrealisierung stattfindet. Erfolgt die Übertragung entgeltlich, so sind Veräußerungstatbestände gem. §§17 und 23 EStG denkbar, jedoch eher als Ausnahmefall.⁷⁹

Keine Gewinnrealisierung tritt außerdem ein, wenn der Stifter nach §6 Abs. 3 EStG seinen Betrieb, einen Teilbetrieb oder seinen Anteil als Mitunternehmer unentgeltlich an die Stiftung weitergibt und somit eine Dotierung aus dem Betriebsvermögen des Stifters stattfindet. Dabei muss die Stiftung eine Buchwertfortführung vornehmen. Es besteht kein Wahlrecht zur Aufdeckung stiller Reserven. Mit den Buchwerten gehen auch die gebildeten Rücklagen nach §§6b bis 6d EStG auf die Stiftung als Rechtsnachfolger über.⁸⁰

Eine Übertragung eines Anteils des Mitunternehmeranteils ist nicht möglich, da diese gesetzlich nur auf eine natürliche Person zulässig ist.⁸¹

Überträgt der Stifter dagegen nur einzelne Gegenstände des Betriebsvermögens gegen Entgelt, führt dies gem. §6 Abs. 1 Nr. 4 EStG zur Realisierung eines steuerpflichtigen Gewinns. Dieser ergibt sich aus der Differenz zwischen dem letzten Wertansatz beim Zuwendenden, also dem bisherigen steuerlichen Buchwert auf Grundlage der Anschaffungs- und Herstellungskosten und dem Wertansatz beim Empfänger in Höhe des Verkehrswertes des Vermögens.⁸² Gleiches gilt, wenn der Stifter im Rahmen der unentgeltlichen Übertragung des Betriebsvermögens einzelne Wirtschaftsgüter zurückhält und diese in sein Privatvermögen überführt.⁸³

Die Realisierung stiller Reserven führt also zur steuerlichen Belastung durch Einkommensteuer und Körperschaftsteuer beim Veräußerer des Vermögens.

Neben der Erbschaft- und Schenkungsteuer fällt unter Umständen auch Umsatzsteuer an, soweit der Stifter ein Unternehmer ist und die Übertragung für ihn einen umsatzsteuerbaren Vorgang darstellt. Zu beachten ist hierbei allerdings §1 Abs. 1a UStG, wonach die Umsätze im Rahmen einer Geschäftsveräußerung nicht der Umsatzsteuer unterliegen.

Enthält das übertragene Vermögen inländische Immobilien, so ist die Übertragung zwar Grunderwerbsteuerbar, da es sich aber meist gleichzeitig um einen erbschaft- und schenkungsteuerbaren Vorgang handelt, entfällt die Grunderwerbsteuer gem. §3 Nr. 2 GrEStG.⁸⁴ Wird die Übertragung entgeltlich vorgenommen, entsteht eine Grunderwerbsteuerpflicht.

3.2. Laufende Besteuerung

Die laufende Besteuerung von privatnützigen Stiftungen weicht gegenüber der Besteuerung von sonstigen Körperschaften kaum ab. Es sind nur wenige stiftungsspezifische Besonderheiten vorhanden.⁸⁵

⁷⁹Vgl. Götz in Berndt and Götz (2009), Rz. 659.

⁸⁰Vgl. Götz in Berndt and Götz (2009), Rz. 733.

⁸¹Vgl. Götz in Berndt and Götz (2009), Rz. 662.

⁸²Vgl. Götz in Berndt and Götz (2009), Rz. 664.

⁸³Vgl. Götz in Berndt and Götz (2009), Rz. 665.

⁸⁴Vgl. Richter, Gollan in Meyn et al. (2013), Rz. 590.

⁸⁵Vgl. Richter, Gollan in Meyn et al. (2013), Rz. 916.

⁷⁴Vgl. G. and T. (2014), S. 79.

⁷⁵Vgl. G. and T. (2014), S. 122.

⁷⁶Vgl. Richter, Gollan in Meyn et al. (2013), Rz. 572.

⁷⁷Vgl. Richter, Gollan in Meyn et al. (2013), Rz. 585.

⁷⁸Vgl. Götz in Berndt and Götz (2009), Rz. 658.

Privatnützige Stiftungen sind während ihres Bestehens nach §1 Abs. 1 Nr. 4 KStG grundsätzlich unbeschränkt körperschaftsteuerpflichtig, da sie zu den juristischen Personen des privaten Rechts zählen. Dabei wird ihr zu versteuern-des Einkommen laut §7 Abs. 1 und §8 Abs. 1 KStG i. V. m. §§2 ff. EStG mit der Körperschaftsteuer und gem. §2 Nr. 3 und §Nr. 1 und 4 SolZG mit dem Solidaritätszuschlag belastet. Grundsätzlich können Stiftungen Einkünfte in allen Einkunftsarten des §2 Abs. 1 EStG erzielen. Dabei sind allerdings Einkünfte aus nichtselbstständiger und selbstständiger Arbeit nur in Ausnahmefällen denkbar.⁸⁶ Unternehmensverbundene Stiftungen erzielen ihr Einkommen meist aus Gewinnen einer gewerblichen Betätigung oder aus Gewinnanteilen einer Mitunternehmerschaft. Sie sind den Einkünften aus einem Gewerbebetrieb nach §15 EStG zuzuordnen. Ist die Stiftung nach den Vorschriften des HGB zur Führung von Büchern verpflichtet, so sind alle Einkünfte der Stiftung den Einkünften aus dem Gewerbebetrieb zuzuordnen.⁸⁷ Erzielt die unternehmensverbundene Stiftung Kapitalerträge i. S. d. §20 Abs. 1 Nr. 1, 2, 9 und 10 Buchst. a EStG, so sind diese i. V. m. §8b Abs. 1 KStG im Ergebnis zu 95 Prozent steuerfrei. 5 Prozent der Bezüge werden gem. §8b Abs. 3 und 5 KStG pauschal als nichtabzugsfähige Betriebsausgaben angesehen und unterliegen damit der Besteuerung. Bei der Einkommensermittlung können außerdem nach §10 Abs. 1 KStG die Aufwendungen für die Erfüllung der Stiftungszwecke nicht abgezogen werden, soweit sie sich der Art nach als Einkommensverwendung darstellen. Aufwendungen, die dem Charakter nach Betriebsausgaben sind, sind nach allgemeinen Gewinnermittlungsgrundsätzen abziehbar.⁸⁸ Der unternehmensverbundenen Stiftung steht gem. §24 Abs. 1 KStG ein Freibetrag i. H. v. 5.000 € zu, der das zu versteuernde Einkommen mindert.

Unterhält die privatnützige Stiftung einen wirtschaftlichen Geschäftsbetrieb, so ist sie gem. §2 Abs. 3 GewStG gewerbsteuerpflichtig. Nur bei Stiftungen, die lediglich Einkünfte aus Kapitalvermögen oder aus Land- und Forstwirtschaft ausweisen, besteht keine Gewerbsteuerpflicht.⁸⁹ Buchführungspflichtige Stiftungen sind schon nach §8 Abs. 2 KStG gewerbsteuerpflichtig.

Gem. §§6 und 7 GewStG wird der nach körperschaftsteuerlichen Vorgaben ermittelte Gewinn, der sich um die Hinzurechnungen gem. §8 GewStG vermehrt und um die Kürzungen gem. §9 GewStG vermindert, mit der Gewerbsteuer belastet.

Die Anteile an Gewinnen bzw. Verlusten aus Mitunternehmerschaften können nach §§8 und 9 Nr. 2 GewStG ggf. bei der Stiftung gewerbsteuerlich erfasst werden.⁹⁰

Ist die Stiftung an einer Kapitalgesellschaft beteiligt und beträgt diese Beteiligung mehr als 15 Prozent des Grundkapitals, wird das Schachtelprivileg nach §9 Nr. 2a GewStG

angewandt. Der Gewinn aus der Veräußerung von Kapitalgesellschaftsanteilen bleibt zu 95 Prozent gewerbsteuerfrei. Veräußerungsverluste können jedoch nicht abgezogen werden.⁹¹

Wie bei der Körperschaftsteuer steht der unternehmensverbundenen Stiftung auch bei der Gewerbsteuer gem. §11 Abs. 1 Satz 3 Nr. 2 GewStG ein Freibetrag von 5.000 € zu, der das Einkommen mindert.

Zur Belastung der Stiftung mit Erbschaft- und Schenkungsteuer kommt es ggf. dann, wenn der Stifter weitere Zuwendungen leistet. Anders als bei Stiftungserrichtung werden hier jedoch keine Steuerklassenvergünstigungen des §15 Abs. 2 Satz 1 ErbStG gewährt. Die Zuwendungen der letzten 10 Jahre an die Stiftung werden nach §14 ErbStG zusammengerechnet. Die daraus resultierende Erbschaftsteuer darf nach §14 Abs. 2 ErbStG 50 Prozent nicht übersteigen.⁹² Besondere Aufmerksamkeit ist bei Umwandlung einer Familienstiftung in eine gewöhnliche Stiftung erforderlich. Diese Umwandlung wird von der Finanzverwaltung als Neugründung einer Stiftung gesehen und wird daher gem. §7 Abs. 1 Nr. 9 i. V. m. §15 Abs. 1 ErbStG nach Steuerklasse III besteuert werden.⁹³ Nur bei Umwandlung einer Familienstiftung in eine gemeinnützige Stiftung erfolgt diese gem. §13 Abs. 1 Nr. 16b ErbStG erbschaftsteuerfrei.

Eine weitere Besonderheit erfährt die Familienstiftung durch die Besteuerung mit der Erbersatzsteuer nach §1 Abs. 1 Nr. 4 ErbStG. Alle 30 Jahre unterliegt das Vermögen einer Familienstiftung der Erbersatzsteuer. Damit soll verhindert werden, dass das Vermögen auf Dauer der Erbschaftsteuer entzogen wird, weil es im Stiftungsvermögen der Familienstiftung gebunden ist. Steuerschuldner ist gem. §20 Abs. 1 ErbStG die Stiftung. Dabei wird der fiktive Fall angenommen, dass alle 30 Jahre das Vermögen auf die nächste Generation, die aus zwei Kindern besteht, übergeht.⁹⁴ Hieraus ergibt sich für die Ermittlung der Besteuerung die Steuerklasse I. Zusätzlich wird gem. §15 Abs. 2 Satz 3 ErbStG der doppelte Freibetrag nach §16 Abs. 1 Nr. 2 ErbStG gewährt und der Steuersatz gewählt, der bei der Hälfte des steuerpflichtigen Vermögens gelten würde. Auch für die Erbersatzsteuer gelten die Steuervergünstigungen von Betriebsvermögen gem. §§13a, 13b ErbStG.

Grundsätzlich sind die Umsätze einer gewerblich tätigen Stiftung umsatzsteuerpflichtig, wobei sich ggf. aus §4 UStG Steuerbefreiungen für die Stiftung ergeben.

Enthalten etwaige Zustiftungen auch Anteile an einer Personengesellschaft, die über Grundbesitz verfügt, kann es nach §1 Abs. 2a GrEStG zur Belastung mit der Grunderwerbsteuer kommen.

Als Inhaberin von inländischen Grundstücken ist die Stiftung grundsteuerpflichtig gem. §2 Nr. 2 GrStG.

Die Leistung einer unternehmensverbundenen Stiftung an die Destinatäre zählen bei diesen entweder zu den Einkünften

⁸⁶Vgl. Richter, Gollan in Meyn et al. (2013), Rz. 921.

⁸⁷Vgl. Gebel (2002), Rz. 1145.

⁸⁸Vgl. BFH vom 05.06.2003, I R 76/01, BStBl II 2005, S. 305.

⁸⁹Vgl. Götz in Berndt and Götz (2009), Rz. 973 f.

⁹⁰Vgl. Götz in Berndt and Götz (2009), Rz. 978.

⁹¹Vgl. Götz in Berndt and Götz (2009), Rz. 980.

⁹²Vgl. Götz in Berndt and Götz (2009), Rz. 988.

⁹³Vgl. Götz in Berndt and Götz (2009), Rz. 990.

⁹⁴Vgl. Götz in Berndt and Götz (2009), Rz. 993.

ten aus Kapitalvermögen und unterliegen gem. §20 Abs. 1 Nr. 9 EStG i. V. m. §43 Abs. 1 Satz 1 Nr. 7a EStG und §43a Abs. 1 Satz 1 Nr. 1 EStG damit der Kapitalertragsteuer oder sie werden den Einkünften gem. §22 Nr. 1 Satz 2 Hs. 2 EStG zugerechnet.⁹⁵ Letztere unterliegen dem Teileinkünfteverfahren nach §3 Nr. 40 EStG und sind damit gem. §3 Nr. 40d EStG zu 40 Prozent steuerbefreit.⁹⁶ Können die Destinäre der Stiftung unmittelbar oder mittelbar Einfluss auf das Ausschüttungsverhalten der Stiftung nehmen, so sind die Leistungen der Stiftung an die Destinäre als Einkünfte aus Kapitalvermögen anzusehen.⁹⁷ Welche Einkunftsart letztendlich ausschlaggebend ist, falls keine Einflussnahme der Destinäre möglich ist, ist nicht abschließend geklärt, da die Zuwendungen ggf. nicht gleichzusetzen sind mit Gewinnausschüttungen an Anteilseigner einer Kapitalgesellschaft, weil die Destinäre weder Eigentümer noch Mitglieder einer Stiftung sind. In diesem Fall sollte die Stiftung eine Auskunft bei der zuständigen Finanzbehörde einholen und die Kapitalertragsteuer zunächst einbehalten.⁹⁸ Ist der Destinär selbst steuerbegünstigt, wird nach §44a Abs. 7 EStG kein Kapitalertragsteuerabzug durchgeführt.⁹⁹

Zuwendungen an die Destinäre aufgrund von Regelungen in der Satzung sind grundsätzlich schenkungsteuerbefreit, da es sich nicht um eine freigebige Zuwendung i. S. d. §7 Abs. 1 Nr. 1 ErbStG handelt. Erhält der Destinär dagegen eine Ausschüttung von Mitteln ohne vertragliche Grundlage in der Satzung, ist die Freigebigkeit anzunehmen und die Übertragung damit schenkungsteuerpflichtig. Gleiches gilt gem. §7 Abs. 1 Nr. 2 ErbStG für Zuwendungen, die einer Auflage unterliegen.

3.3. Besteuerung der Stiftungsaufhebung

Die Stiftung ist grundsätzlich auf Dauer angelegt. Nur unter bestimmten Voraussetzungen können Stiftungen aufgelöst werden, sei es bspw. durch in der Stiftungssatzung geregelte Ausnahmen oder durch staatliche Auflösung aufgrund einer Gemeinwohlgefährdung gem. §87 Abs. 1 BGB.¹⁰⁰

Steuerlich ist die Stiftungsaufhebung vergleichbar mit der Besteuerung bei Stiftungserrichtung – allerdings in umgekehrter Weise.¹⁰¹

In der Regel fällt im Rahmen der Stiftungsaufhebung das Stiftungsvermögen an in der Stiftungssatzung bestimmte Personen. Fehlt eine solche Regelung in der Satzung, so ist letztendlich der Staat Anfallberechtigter. Die Auflösung einer Stiftung wird als Liquidation angesehen. Die Auflösung nach §11 KStG findet auf Stiftungen keine Anwendung. Deshalb erfolgt die Besteuerung gem. §8 KStG nach §§16 und 17 EStG.¹⁰² Sind die Destinäre der Stiftung gem. Satzung

Anfallsberechtigte und werden ihnen Betriebsvermögensteile ohne Entgeltzahlung überlassen, so ist gem. §16 Abs. 3 Satz 3 EStG der gemeine Wert anzusetzen. Bei der Übertragung von Betrieb, Teilbetrieb oder Mitunternehmerschaft können auch die Buchwerte übernommen werden, vorausgesetzt die Anforderungen aus §6 Abs. 3 EStG werden erfüllt. Gehen die Betriebsvermögensteile nicht an die Destinäre über, so sind die stillen Reserven besteuernspflichtig. Dabei können keine Freibeträge angerechnet und keine Steuerermäßigung bzw. kein Steuernachlass gewährt werden.¹⁰³

Die Gewerbesteuerpflicht der Stiftung endet mit der tatsächlichen Aufgabe des Geschäftsbetriebes, die im Falle der Stiftungsauflösung eintritt.¹⁰⁴ Ein daraus entstehender Aufgabegewinn ist gem. §16 Abs. 1 bis 3 EStG gewerbesteuerfrei.

Umsatzsteuer fällt dann an, wenn die Stiftung als Unternehmerin durch die Veräußerung von Gegenständen des Betriebsvermögens Entgelte erwirtschaftet und diese nach den Vorschriften des UStG umsatzsteuerpflichtig sind.¹⁰⁵

Gem. §7 Abs. 1 Nr. 9 ErbStG sind die Vermögenswerte, die bei Aufhebung der Stiftung erworben werden, als Schenkung anzusehen. Schenker ist nach §15 Abs. 2 Satz 2 der Stifter. Fällt das Vermögen der Stiftung direkt an den Stifter zurück, ist dieses in der ungünstigsten Steuerklasse III steuerpflichtig. Daher sollte in der Stiftungssatzung immer eine Vermögensübertragung im Falle der Auflösung an die Abkömmlinge des Stifters vorgesehen werden. Wurde die Stiftung von mehreren Stiftern errichtet, können unterschiedliche Steuerklassen für die Anfallberechtigten angewendet werden.¹⁰⁶

Die Umwandlung einer steuerpflichtigen in eine steuerbefreite Stiftung ist nach §13 KStG körperschaftsteuerbefreit und nach §13 Abs. 1 Nr. 16b ErbStG schenkungsteuerfrei.

4. Steuerbegünstigungen für gemeinnützige unternehmensverbundene Stiftungen

In der Praxis wählen Stifter weitaus häufiger gemeinnützige als rein privatnützige Zwecke, die sie mit ihrer Stiftung fördern möchten: In 2013 verfolgten 95 Prozent aller Stiftungen in Deutschland gemeinnützige Zwecke.¹⁰⁷ Wobei der Begriff der Gemeinnützigkeit hier im weiteren Sinne gelten soll, also sowohl gemeinnützige als auch mildtätige und kirchliche Zwecke beinhaltet.¹⁰⁸ Durch die Verfolgung gemeinnütziger Zwecke erhält die Stiftung umfassende Steuerbegünstigungen oder -befreiungen, sofern die gemeinnützigkeitsrechtlichen Anforderungen erfüllt werden. Gesetzliche Regelungen hierzu finden sich in §§51 bis 68 AO.

⁹⁵Vgl. Gebel (2002), Rz. 1161 i. V. m. Richter, Gollan in Meyn et al. (2013), Rz. 937.

⁹⁶Vgl. Pöllath, Richter in Seifart (2009), § 41 Rz. 62.

⁹⁷Vgl. BFH vom 03.11.2010, I R 98/09, DStR 2011 S. 403 ff.

⁹⁸Vgl. Richter, Gollan in Meyn et al. (2013), Rz. 940.

⁹⁹Vgl. Richter, Gollan in Meyn et al. (2013), Rz. 943.

¹⁰⁰Vgl. Schiffer and Schürmann, Auflösung rechtsfähiger Stiftungen.

¹⁰¹Vgl. Wigand in Wigand et al. (2015), Kap. 5 Rz. 5.

¹⁰²Vgl. Götz in Berndt and Götz (2009); Rz. 1066.

¹⁰³Vgl. Götz in Berndt and Götz (2009); Rz. 1067.

¹⁰⁴Vgl. Götz in Berndt and Götz (2009), Rz. 1071.

¹⁰⁵Vgl. Götz in Berndt and Götz (2009), Rz. 1072.

¹⁰⁶Vgl. Götz in Berndt and Götz (2009), Rz. 1081.

¹⁰⁷Vgl. Bundesverband deutscher Stiftungen (2014), S. 25.

¹⁰⁸Vgl. Schlüter and Stolte (2016), Kap. 7 Rz. 11.

4.1. Gemeinnützigkeitsrechtliche Anforderungen für Steuerbefreiungen

Die Voraussetzungen zur Steuerbefreiung bei Stiftungen lassen sich in zwei Gruppen einteilen. Zum einen geht es um die steuerbegünstigte Verwendung der Stiftungsmittel und zum anderen um die Einkommenserzielung.

Erste Anforderung zur Anerkennung der Gemeinnützigkeit ist gem. §51 AO die ausschließliche Verwendung des Einkommens der Stiftung für gemeinnützige, mildtätige oder kirchliche Zwecke. Diese werden in §§52 bis 54 AO näher erläutert und müssen immer anhand des konkreten Einzelfalls von der zuständigen Behörde vor Anerkennung geprüft werden.¹⁰⁹ Neben der Einkommensverwendung für bestimmte Zwecke spielt auch die Umsetzung eine Rolle. Bei der Verwendung des Einkommens muss die Stiftung die Grundsätze der Selbstlosigkeit, der Ausschließlichkeit und der Unmittelbarkeit erfüllen. Zusätzlich werden weitere Anforderungen an die Satzung und die Geschäftsführung gestellt.

Neben der Verwendung spielt auch die Einkommenserzielung für die Steuerbefreiung von Stiftungen eine Rolle. Erst bei Einhalten aller Voraussetzungen, die die Einkommenserzielung betreffen, ergibt sich die objektive Steuerbegünstigung.¹¹⁰ Die Tätigkeit der Stiftung darf ausschließlich aus Spenden, Zustiftungen und aus der Verwaltung von Vermögen finanziert sein.¹¹¹ Die Steuervergünstigungen können nicht gewährt werden, wenn die Stiftung Einkommen aus einem wirtschaftlichen Geschäftsbetrieb erzielt. Hiermit soll der Schutz der Wettbewerber gewährleistet werden.¹¹² Anders als bei den Voraussetzungen der Einkommensverwendung und den Anwendungsvoraussetzungen wird bei Verstößen gegen die Voraussetzung für steuerfreie Einkommenserzielung nur jeweils der Teil der Einnahmen versteuert, der den Voraussetzungen nicht entspricht; die Stiftung ist dann also partiell steuerpflichtig.¹¹³

4.1.1. Mittelverwendung für gemeinnützige, mildtätige und kirchliche Zwecke

Die gemeinnützigen Zwecke im engeren Sinne des §52 Abs. 1 AO umfassen alle Zwecke, die das Gemeinwohl oder die Allgemeinheit in materieller, geistiger oder sittlicher Weise selbstlos fördern. Zu diesen zählen bspw. die Förderung von Wissenschaft und Forschung, von Kunst und Kultur oder der Jugend- und Altenhilfe. In §52 Abs. 2 AO findet sich seit der Einführung des „Gesetzes zur weiteren Stärkung des bürgerlichen Engagements“ eine grundsätzlich abschließende Liste aller als gemeinnützig anerkannten Zwecke.¹¹⁴ Zusätzlich können im konkreten Einzelfall weitere Zwecke von der Rechtsprechung und Verwaltung als gemeinnützig anerkannt werden.¹¹⁵ Die gesetzliche Öffnungsklausel nach §52

Abs. 2 Satz 2 AO spielt in der Praxis jedoch kaum eine Rolle, da Zwecke, die schon bei Erstellung des Katalogs bekannt waren, aber bewusst nicht aufgenommen wurden, nicht damit verbunden sind. Nur wenn sich die gesellschaftlichen Verhältnisse verändern, kommt die Öffnungsklausel in Betracht.¹¹⁶

Die Förderung der Allgemeinheit, die als Voraussetzung für die Gemeinnützigkeit festgelegt ist, wird in §52 Abs. 1 AO negativ abgegrenzt: die Allgemeinheit ist keine fest abgeschlossene Gruppe, wie bspw. die Zugehörigkeit zu einer Familie oder zur Belegschaft eines Unternehmens. Des Weiteren handelt es sich nicht um eine dauerhaft sehr kleine Gruppe, die z. B. durch räumliche oder berufliche Merkmale abgegrenzt ist. Auf der anderen Seite bedeutet die Förderung der Allgemeinheit jedoch nicht, dass stets die ganze Allgemeinheit gefördert werden muss. Dies wäre aufgrund der unterschiedlichen Bedürfnisse verschiedener Personen und Personenkreise wohl eine nicht erfüllbare Anforderung.¹¹⁷ Die negative Abgrenzung der Allgemeinheit ist jedoch nicht unumstritten. Eine kleine Stiftung wird regelmäßig versuchen, ihren Wirkungskreis so klein zu halten, dass eine spürbare Wirkung in Einzelfällen erreicht werden kann. Damit werden die Mittel der Stiftung im Zweifelsfall nur einer einzigen Person zur Verfügung gestellt, z. B. bei der Vergabe eines Stipendiums. Steuerschädlich wird die Begrenzung auf einen kleinen Personenkreis nur, wenn sie gemeinnützigkeitswidrig ist und die Stiftung nicht selbstlos handelt. Die Konzentration des Mitteleinsatzes, wie im Fall einer Stipendienvergabe, dient hier dem Interesse des gemeinnützigen Zwecks.¹¹⁸

Mildtätige Zwecke nach §53 AO beinhalten zum einen die selbstlose Unterstützung von persönlich hilfsbedürftigen Personen aufgrund von körperlichen, geistigen oder seelischen Zuständen. Zum anderen ist die Unterstützung von wirtschaftlich bedürftigen Personen mildtätig, deren allgemeine Bezüge nicht über dem Vierfachen des Regelsatzes der Sozialhilfe liegen und die ihren Unterhalt nicht nachhaltig aus ihrem Vermögen verbessern können. Auch andere besondere Gründe wie bspw. Naturkatastrophen oder lange Krankheiten können zu einer wirtschaftlichen Notlage führen.¹¹⁹ Im Unterschied zu den gemeinnützigen Zwecken dürfen mildtätige Zwecke auf einen dauerhaft begrenzten Personenkreis beschränkt sein. Gesetzlich ist keine Förderung der Allgemeinheit vorgegeben. Lediglich die selbstlose Hilfe muss als Hauptmotiv erhalten bleiben. Die Förderung von ausschließlich verwandten Hilfsbedürftigen ist nicht als mildtätig einzustufen, da hier die Förderung der Verwandtschaft im Vordergrund steht.¹²⁰

Kirchliche Zwecke nach §54 AO werden dann von einer Stiftung verfolgt, wenn ihre Tätigkeit darauf gerichtet ist, eine Religionsgemeinschaft des öffentlichen Rechts selbstlos zu fördern. Dazu zählen insbesondere die katholische und

¹⁰⁹Vgl. Wigand in Wigand et al. (2015), Kap. 5 Rz. 38.

¹¹⁰Vgl. Fleschutz (2008), S. 96.

¹¹¹Vgl. Schlüter and Stolte (2016), Kap. 7 Rz. 2.

¹¹²Vgl. Pöllath, Richter in Seifart (2009), § 43 Rz. 9.

¹¹³Vgl. Schlüter and Stolte (2016), Kap. 7 Rz. 129.

¹¹⁴Vgl. Wigand in Wigand et al. (2015), Kap. 5 Rz. 40.

¹¹⁵Vgl. Wigand in Wigand et al. (2015), Kap. 5 Rz. 41 f.

¹¹⁶Vgl. OFD Koblenz vom 11.01.2010, S 0171 A – St 331, DB 2010, S. 756.

¹¹⁷Vgl. Pöllath, Richter in Seifart (2009), § 43 Rz. 31.

¹¹⁸Vgl. Pöllath, Richter in Seifart (2009), § 43 Rz. 32.

¹¹⁹Vgl. Wigand in Wigand et al. (2015), Kap. 5 Rz. 44, so auch Pöllath, Richter in Seifart (2009), § 43 Rz. 39 f.

¹²⁰Vgl. Schlüter and Stolte (2016), Kap. 7 Rz. 38.

evangelische Kirche sowie die jüdischen Kultusgemeinden in Deutschland.¹²¹ Kirchliche Zwecke können nach §54 Abs. 2 AO bspw. die Errichtung, Ausschmückung und Unterhaltung von Gotteshäusern oder die Erteilung von Religionsunterricht sein.

4.1.2. Grundsätze zur Mittelverwendung

Neben der Voraussetzung der Mittelverwendung für gemeinnützige, mildtätige und kirchliche Zwecke müssen auch gewisse Grundsätze bei der Verwendung erfüllt sein: Die Stiftung muss ihre Zwecke selbstlos, ausschließlich und unmittelbar i. S. d. §§55 bis 57 AO verfolgen.

Verfolgt eine Stiftung in erster Linie gewerbliche oder sonstige Erwerbszwecke, handelt sie nicht selbstlos und damit gemeinnützigkeitsschädlich. Das bedeutet, dass die Tätigkeit der Stiftung immer auf den gemeinnützigen Zweck ausgerichtet sein muss, damit die Selbstlosigkeit nach §55 AO gewährleistet ist. Eine wirtschaftliche Betätigung ist ausschließlich dann selbstlos, wenn sie lediglich Mittel zum steuerbegünstigten Zweck ist.¹²² Steht die reine Eigennutzung der Mitglieder einer Stiftung im Vordergrund, so ist die Selbstlosigkeit gefährdet. Eigenwirtschaftliche Zwecke sind nur als untergeordnete Zwecke erlaubt. Der gemeinnützige Zweck der Stiftung muss die Betätigung beherrschen.¹²³ Als Kriterium zur Überprüfung der Stellung der verschiedenen Zwecke in ihrer Verfolgung, soll der zeitliche und personelle Aufwand der steuerbegünstigten Zwecke dem Aufwand der wirtschaftlichen und damit steuerpflichtigen Zwecke gegenübergestellt werden.

Zu der selbstlosen Tätigkeit zählt auch die zeitnahe Verwendung der finanziellen Mittel der Stiftung. Zeitnah bedeutet, dass die Mittel, die der Stiftung zugeflossen sind, gem. §55 Abs. 1 Nr. 5 Satz 3 AO spätestens 24 Monate nach dem Ende des Kalender- oder Wirtschaftsjahres zur Erfüllung der satzungsmäßigen Zwecke verwendet werden müssen.¹²⁴ Nur für einige in §62 AO gesetzlich geregelte Fälle ist es der Stiftung gestattet, Rücklagen zu bilden. Sie sind als Ausnahme der sonst geltenden zeitnahen Mittelverwendung anzusehen. Zu den Mitteln gem. §55 AO zählen nicht nur zur Verfügung stehende Geldbeträge aus Spenden, Beiträgen, Erträgen des Vermögens und dem begünstigten Zweckbetrieb, sondern auch die Gewinne aus dem steuerpflichtigen wirtschaftlichen Geschäftsbetrieb.¹²⁵ Ausgenommen vom Gebot der Mittelverwendung ist das sog. Stiftungsvermögen, die zulässigen Rücklagen, die die Stiftung gebildet hat und sonstige Einlagen, die vom Stifter, von Gesellschaftern oder anderen Mitgliedern nach §62 Abs. 3 AO zur Erhöhung des Grundstockvermögens geleistet wurden.¹²⁶ Des Weiteren besteht gem. §55 Abs. 1 Nr. 1 Satz 2 AO ein Zuwendungsverbot der Stiftung an ihren Stifter, dessen Erben, Angehörige

oder sonstige nahe stehende Personen. Bei den Zuwendungen kann es sich hierbei um sämtliche wirtschaftliche Vorteile handeln. Zu unterscheiden sind diese Zuwendungen von der Vergütung, die den Organmitgliedern für ihre Tätigkeiten gezahlt werden, die allerdings in der Satzung detailliert festgehalten werden müssen.¹²⁷ Die Grenze der Vergütung liegt in der Verhältnismäßigkeit. Unverhältnismäßig hohe Zahlungen an die Organmitglieder sind gemeinnützigkeitsschädlich. Dabei muss die Höhe der erwarteten Erträge der Stiftung berücksichtigt werden.¹²⁸

Nach §55 Abs. 1 Nr. 4 AO gilt der Grundsatz der Vermögensbindung. Im Falle der Auflösung der Stiftung oder beim Wegfall des gemeinnützigen Zwecks muss das Vermögen endgültig für steuerbegünstigte Zwecke verwendet werden – entweder von der Stiftung selbst oder nach Übertragung von einer anderen steuerbegünstigten Körperschaft.

In §62 Abs. 1 AO sind einige Ausnahmen des Gebots der zeitnahen Mittelverwendung enthalten. Es handelt sich jedoch nicht um eine abschließende Aufzählung. Die Bildung einer gebundenen Rücklage gem. §62 Abs. 1 Nr. 1 AO ist zulässig, wenn die steuerbegünstigten Zwecke von der Stiftung ausnahmsweise nicht ohne eine vorübergehende Ansammlung von Mitteln nachhaltig erfüllt werden kann. Dabei muss die Stiftung allerdings bereits im Voraus festlegen, zu welchem Zeitpunkt und wofür sie die Mittel der Rücklage genau verwendet. Erfordert die Finanzierung eines Projektes die Mittel aus Rücklagen mehrerer Jahre, so ist zusätzlich ein konkreter Zeit- und Budgetplan zur Projektdurchführung aufzustellen. Die Rücklagen hierfür werden als Projektrücklagen bezeichnet.¹²⁹ Eine weitere Form der gebundenen Rücklage ist die Wiederbeschaffungsrücklage nach §62 Abs. 1 Nr. 2 AO. Sie kann von der Stiftung genutzt werden, um Mittel anzusparen, die der tatsächlich geplanten Wiederbeschaffung eines für die satzungsmäßigen Zwecke genutzten Wirtschaftsgutes dienen. Dazu zählt auch die sog. Investitionsrücklage zur Finanzierung eines Immobilienneubaus o. Ä.¹³⁰ Die letzte Gruppe der gebundenen Rücklagen umfasst die Rücklage für Betriebsmittel, für den Fall, dass eine Stiftung auf eine Reserve in Notfällen für laufende satzungsmäßige Aufwendungen zurückgreifen muss, falls die Einnahmen der Stiftung vorübergehend nicht ausreichen.¹³¹ Eine Stiftung kann außerdem eine freie Rücklage nach §62 Abs. 1 Nr. 3 AO bilden, um die Leistungsfähigkeit zu gewährleisten. Allerdings führt das aktuell niedrige Zinsniveau häufig dazu, dass die Einnahmequelle zur Gewährleistung der Leistungsfähigkeit nicht mehr ausreicht. Zulässig ist die Bildung von Rücklagen bis zu einem Drittel des Überschusses der Einnahmen über die Vermögensverwaltungskosten.¹³² Nach §62 Abs. 1 Nr. 4 AO können Stiftungen Rücklagen bilden, um Beteiligungen an Kapitalgesellschaften zu erwerben. Insbesondere Be-

¹²¹Vgl. Schlüter and Stolte (2016), Kap. 7 Rz. 40 f.

¹²²Vgl. Schlüter and Stolte (2016), Kap. 7 Rz. 44.

¹²³Vgl. Pöllath, Richter in Seifart (2009), § 43 Rz. 44 und 46.

¹²⁴Vgl. Schlüter and Stolte (2016), Kap. 7 Rz. 61.

¹²⁵Vgl. Pauli in Stumpf et al. (2011), Teil E Kap. 2 § 55 AO Rz. 11.

¹²⁶Vgl. Pauli in Stumpf et al. (2011), Teil E Kap. 2 § 55 AO Rz. 12.

¹²⁷Vgl. Pauli in Stumpf et al. (2011), Teil E Kap. 2 § 55 AO Rz. 14 i. V. m. Schlüter and Stolte (2016), Kap. 2 Rz. 70.

¹²⁸Vgl. Meyn in Meyn et al. (2013), Rz. 181 f.

¹²⁹Vgl. Schlüter and Stolte (2016), Kap. 7 Rz. 70.

¹³⁰Vgl. Schlüter and Stolte (2016), Kap. 7 Rz. 71.

¹³¹Vgl. Schlüter and Stolte (2016), Kap. 7 Rz. 72.

¹³²Vgl. Schlüter and Stolte (2016), Kap. 7 Rz. 73.

teiligungsträgerstiftungen können davon Gebrauch machen, eine Ansparrücklage zu bilden, um Anteile an dem gehaltenen Unternehmen zu übernehmen, mit dem Ziel, die prozentuale Beteiligungshöhe und etwaige Mehrheitsrechte zu erhöhen oder beizubehalten, wenn das Unternehmen eine Kapitalerhöhung durchführt.¹³³ Neben den genannten Rücklagen ist es Stiftungen gestattet, Umschichtungsrücklagen und Rücklagen im wirtschaftlichen Geschäftsbetrieb und im Bereich der Vermögensverwaltung zu bilden. Auf diese soll jedoch nicht genauer eingegangen werden. Zudem gilt ein weitgehendes Rücklagenverbot und Thesaurierungsverbot.

Neben der selbstlosen Tätigkeit einer Stiftung muss diese auch ausschließlich und unmittelbar erfolgen. Nach §56 AO ist die Stiftung dazu verpflichtet, ihre Tätigkeit auf die Verfolgung der satzungsmäßig festgelegten Zwecke zu konzentrieren. Sollen weitere gemeinnützige Zwecke verfolgt werden, so muss dies in der Satzung entsprechend geändert oder ergänzt werden, da sonst die verwendeten Mittel steuerschädlich genutzt werden.¹³⁴

Eine Stiftung handelt unmittelbar i. S. d. §57 AO, wenn sie bei der Verfolgung des Stiftungszwecks selbst tätig wird. Dies kann durch ihre Organe und Vertreter geschehen oder durch Hilfspersonen, soweit deren Handeln nach Weisung der Stiftung geschieht.¹³⁵ Diese Weisungsabhängigkeit wird durch Verträge oder organisatorische Einbindung gegenüber der Stiftung erreicht.¹³⁶ Eine gemeinnützige Stiftung kann auch durch fördernde Tätigkeiten ihre Zwecke verwirklichen. Dabei gibt die Stiftung ihre Mittel an andere gemeinnützige Körperschaften weiter.¹³⁷ Die Unterstützung kann nach §58 Nr. 1, 2 und 3 AO durch Geld- und Sachzuwendungen, nach §58 Nr. 4 AO durch eine unentgeltliche Überlassung von Personal oder nach §58 Nr. 5 AO durch die unentgeltliche Überlassung von Räumlichkeiten erfolgen. Eine Stiftung, die solche fördernden Zwecke vorsieht, kann sich ausschließlich auf diese beschränken. Ist die Förderung anderer gemeinnütziger Körperschaften jedoch nicht im Stiftungszweck festgehalten, dürfen die Ausgaben der Förderung anderer gemeinnütziger Körperschaften nicht mehr als die Hälfte der Gesamtausgaben betragen.¹³⁸

4.1.3. Formelle Voraussetzungen

Neben den materiellen Voraussetzungen aus §§52 bis 58 AO sind unter §§59 bis 61, 63 AO Regelungen angegeben, unter welchen formellen Voraussetzungen die Steuerbegünstigung gewährt wird.¹³⁹ Dabei werden besondere Anforderungen an die formelle Satzungsmäßigkeit im Sinne einer dem Gesetz entsprechenden Satzung und an die materielle

Satzungsmäßigkeit im Sinne der tatsächlichen Geschäftsführung gestellt.¹⁴⁰

Nach §60 AO muss aus der Satzung einer steuerbegünstigten Körperschaft hervorgehen, dass diese einem oder mehreren gemeinnützig anerkannten Zwecken selbstlos, ausschließlich und unmittelbar nachgeht und damit die satzungsmäßigen Voraussetzungen für die Steuervergünstigung erfüllt werden.¹⁴¹ Die erforderlichen Angaben müssen in der Satzung selbst vorhanden sein; es ist nicht ausreichend, auf andere Schriftstücke oder gesetzliche Regelungen zu verweisen.¹⁴² Es bietet sich für den Stifter an, sich nah an die Mustersatzung zu halten, die dem Anwendungserlass zur AO beigefügt ist, um die Steuerbegünstigung nicht zu gefährden. Trotzdem ist es unbedingt empfehlenswert, die Satzung vor rechtlicher Anerkennung durch die Stiftungsaufsicht vom zuständigen Finanzamt prüfen zu lassen. Eine folgende Anerkennung des Finanzamtes ist zwar nicht rechtskräftig, sie bietet jedoch einen gewissen faktischen Schutz.¹⁴³ Die exakte Formulierung der Satzung sollte mit besonderer Sorgfalt durchgeführt werden. Dies gilt insbesondere für die zu verfolgenden Zwecke und die Mittel und Instrumente, die dafür aufgewandt werden. Es bietet sich bspw. an, jegliche Tätigkeiten, die der Vermögensverwaltung oder einen wirtschaftlichen Bereich der Stiftung betreffen, nicht unter der Thematik des Stiftungszwecks, sondern an anderer Stelle in der Satzung zu erläutern.¹⁴⁴ Nach §59 AO ist neben der Notwendigkeit der Angabe eines Stiftungszwecks auch eine Darlegung erforderlich, mit welchen Maßnahmen und durch welche Tätigkeiten im Einzelnen die genannten Zwecke erfüllt werden.¹⁴⁵

Im Sinne der Vermögensbindung muss das Vermögen bei Auflösung oder Aufhebung der Stiftung einem gemeinnützigen Zweck gewidmet sein. Dieser Zweck muss ferner nach §61 Abs. 1 AO als Teil der sog. Anfallklausel in der Satzung aufgenommen werden, sodass auf Grund derer die Steuerbegünstigung geprüft werden kann.¹⁴⁶ Es ergeben sich für die Angabe des künftigen Verwendungszwecks zwei Möglichkeiten: Als Empfänger des Stiftungsvermögens kann eine bestimmte juristische Person des öffentlichen Rechts bzw. eine steuerbegünstigte Körperschaft benannt werden oder die Satzung beschränkt sich auf allgemein gehaltene Angaben zu der begünstigten Körperschaft, wobei in diesem Fall der steuerbegünstigte Zweck genau genannt werden muss. Eine Stiftungssatzung die vorsieht, das Vermögen an eine nicht näher bestimmte gemeinnützige Körperschaft ohne konkreten Verwendungszweck zu übertragen, erfüllt nicht die Anforderungen des §61 AO.¹⁴⁷ Bei Betrachtung der möglichen Folgen bei Verstößen gegen das Gebot der Vermögensbindung wird

¹³³Vgl. Schlüter and Stolte (2016), Kap. 7 Rz. 77.

¹³⁴Vgl. Richter, Gollan in Meyn et al. (2013), Rz. 392 f.

¹³⁵Vgl. Richter, Gollan in Meyn et al. (2013), Rz. 405 bis 407.

¹³⁶Vgl. Wigand in Wigand et al. (2015), Kap. 5 Rz. 52.

¹³⁷Vgl. Wigand in Wigand et al. (2015), Kap. 5 Rz. 54.

¹³⁸Vgl. OFD Frankfurt am Main vom 21.01.2008, S 0177 A – 6 – St 53/ S 0177 A – 7 – St 53, DStR 2008, S. 406, 407.

¹³⁹Vgl. Pöllath, Richter in Seifart (2009), § 43 Rz. 77.

¹⁴⁰Vgl. Pöllath, Richter in Seifart (2009), § 43 Rz. 78.

¹⁴¹Vgl. Richter, Gollan in Meyn et al. (2013), Rz. 428.

¹⁴²Vgl. BFH vom 05.08.1992, X R 165/88, BStBl II 1992, S. 1048.

¹⁴³Vgl. Pöllath, Richter in Seifart (2009), § 43 Rz. 83.

¹⁴⁴Vgl. Richter, Gollan in Meyn et al. (2013), Rz. 432.

¹⁴⁵Vgl. Richter, Gollan in Meyn et al. (2013), Rz. 434.

¹⁴⁶Vgl. Pauli in Stumpf et al. (2011), Teil E Kap. 2 § 55 AO Rz. 23 i. V. m. Schlüter and Stolte (2016), Kap. 7 Rz. 60.

¹⁴⁷Vgl. Richter, Gollan in Meyn et al. (2013), Rz. 438 f.

deutlich, wie bedeutsam eine korrekte Gestaltung der Anfallklausel ist. Wird die satzungsmäßige Vermögensbindung verletzt oder aufgehoben, so kommt es zur rückwirkenden Aufhebung der Steuervergünstigung. Dabei kommt es u. a. zur nachträglichen Besteuerung der Überschüsse aus der Vermögensverwaltung, zur nachträglichen Besteuerung der Gewinne aus Zweckbetrieben, ggf. zur nachträglichen Steuerbelastung von Spenden und Zustiftungen und zum Wegfall der Berechtigung zum Empfang steuerlich abzugsfähiger Spenden.¹⁴⁸

Nach §63 Abs. 1 AO muss die tatsächliche Geschäftsführung auf die ausschließliche und unmittelbare Erfüllung der steuerbegünstigten Zwecke gerichtet sein. Grundlage der tatsächlichen Geschäftsführung sind dabei die Satzungsbestimmungen.¹⁴⁹ Der Stiftungsvorstand trägt nach §63 Abs. 3 AO die Nachweispflicht und muss deshalb Bücher über die Einnahmen und Ausgaben der Stiftung führen und dem Finanzamt die Bildung und Entwicklung der Rückstellungen nachweisen.¹⁵⁰ Die tatsächliche Geschäftsführung verstößt insbesondere dann gegen gemeinnützigkeitsrechtliche Anforderungen, wenn ihre Aktivitäten nicht ausschließlich auf steuerbegünstigte Zwecke ausgerichtet sind, oder sie die Mittel nicht zeitnah für steuerbegünstigte Zwecke nutzt.¹⁵¹ Ein endgültiges Erreichen des Zwecks ist aber nicht unbedingt erforderlich. Des Weiteren sind auch vorbereitende Tätigkeiten, die der Umsetzung der Satzungszwecke vorausgehen, zulässig.¹⁵² Bei einem Verstoß gegen die gemeinnützigkeitsrechtlichen Voraussetzungen kann es letztendlich dazu führen, dass die Finanzverwaltung die Gemeinnützigkeit entzieht. Zuvor werden jedoch i. d. R. unter Beachtung des Grundsatzes der Verhältnismäßigkeit mildere Maßnahmen gewählt, wie bspw. eine geeignete Fristsetzung zur zeitnahen Mittelverwendung nach §63 Abs. 4 AO.

Wird eine der materiellen oder formellen Voraussetzungen nicht erfüllt, so kann insgesamt keine Steuervergünstigung gewährt werden.¹⁵³ Bei Erfüllung aller Voraussetzungen gilt die subjektive Steuerbegünstigung als gegeben.¹⁵⁴

4.2. Besteuerung gemeinnützig anerkannter Stiftungen und ihrer wirtschaftlichen Tätigkeit

Vor der Gründung einer Stiftung ist es empfehlenswert, eine Vorabstimmung mit dem zuständigen Finanzamt zur Anerkennung der Gemeinnützigkeit durchzuführen.¹⁵⁵ Damit es dann letztendlich zur Steuerbefreiung der Stiftung kommt, sind zwei Bestätigungen des Finanzamtes notwendig. Zum einen braucht es eine vorläufige Bescheinigung nach §60a AO, die die satzungsmäßigen Voraussetzungen feststellt, und zum anderen den Freistellungsbescheid, der

die formalen Voraussetzungen bestätigt. Da die formalen Anforderungen nicht nur die Ansprüche an die Satzung beinhalten, sondern auch die tatsächliche Geschäftsführung im Veranlagungszeitraum inkludiert wird, ist die Prüfung dieser Voraussetzungen nur rückwirkend möglich.¹⁵⁶ Die vorläufige Bescheinigung ermöglicht der Stiftung daher zunächst ihre Handlungsfähigkeit z. B. im Hinblick auf die Ausstellung von Zuwendungsbestätigungen. Der Freistellungsbescheid wird nach Einreichen der Steuererklärung erteilt, solange alle Voraussetzungen erfüllt sind. Andernfalls erhält die Stiftung einen Steuerbescheid und ist damit steuerpflichtig.¹⁵⁷

Im Rahmen der Besteuerung bei Errichtung der Stiftung entfällt die Erbschaft- bzw. Schenkungsteuer bei Vermögensübertragung an die gemeinnützige Stiftung gem. §13 Abs. 1 Nr. 16b ErbStG, wenn diese nach ihrer Satzung und ihrer tatsächlichen Geschäftsführung im Zeitpunkt der Zuwendung und für mindestens 10 Jahre nach der Zuwendung mit ihrem Vermögen gemeinnützige Zwecke fördert. Zusätzlich muss die formelle Satzungsmäßigkeit nach §§59 bis 61 AO erfüllt sein.

Bringt der Stifter seinen Betrieb, einen Teilbetrieb oder einen Mitunternehmeranteil unentgeltlich in die steuerbegünstigte Stiftung ein und entsteht dadurch ein steuerpflichtiger wirtschaftlicher Geschäftsbetrieb, erfolgt die Übertragung nach §6 Abs. 4 Satz 4 EStG steuerneutral zum Buchwert und es werden keine stillen Reserven aufgedeckt, die beim Stifter zur Besteuerung führen könnten.¹⁵⁸

Die steuerlichen Vergünstigungen für gemeinnützige Stiftungen zählen in erster Linie für den Tätigkeitsbereich der Stiftung, der sich auf die Gemeinnützigkeit bezieht. Deshalb ist es im Zuge der laufenden Besteuerung notwendig, alle Einnahmen und Ausgaben der Stiftung einem bestimmten Tätigkeitsbereich zuzuordnen. Nach dem sog. Vier-Sphären-Modell werden gemeinnützigkeitsrechtlich und steuerlich vier Bereiche unterschieden. Die Tätigkeiten einer Stiftung werden sodann entweder dem ideellen Bereich, dem steuerfreien wirtschaftlichen Geschäftsbetrieb (sog. Zweckbetrieb), dem steuerpflichtigen wirtschaftlichen Geschäftsbetrieb oder der Vermögensverwaltung zugeordnet.¹⁵⁹

Der ideelle Bereich der Stiftungstätigkeit besteht zum einen aus der Mittelverwendung für den in der Stiftungssatzung festgelegten Zweck und zum anderen aus Einnahmen aus steuerbegünstigten Zuwendungen wie Spenden und Zustiftungen. Einkünfte aus der laufenden Tätigkeit innerhalb des ideellen Bereiches sind gem. §5 Abs. 1 Nr. 9 KStG von der Körperschaftsteuer und gem. §3 Nr. 6 GewStG von der Gewerbesteuer befreit, da die Stiftung hier ohne Gewinnerzielungsabsicht tätig ist und keine wirtschaftlichen Vorteile erzielt werden.¹⁶⁰

Grundsätzlich kann die steuerbegünstigte Stiftung neben ihrem ideellen Bereich auch einer wirtschaftlichen Tätigkeit

¹⁴⁸Vgl. Richter, Gollan in Meyn et al. (2013), Rz. 443 ff.

¹⁴⁹Vgl. Richter, Gollan in Meyn et al. (2013), Rz. 706.

¹⁵⁰Vgl. Wigand in Wigand et al. (2015), Kap. 5 Rz. 90.

¹⁵¹Vgl. Wigand in Wigand et al. (2015), Kap. 5 Rz. 91.

¹⁵²Vgl. Pöllath, Richter in Seifart (2009), § 43 Rz. 92.

¹⁵³Vgl. Pöllath, Richter in Seifart (2009), § 43 Rz. 14.

¹⁵⁴Vgl. Fleschutz (2008), S. 96.

¹⁵⁵Vgl. Schlüter and Stolte (2016), Kap. 7 Rz. 224.

¹⁵⁶Vgl. Schlüter and Stolte (2016), Kap. 7 Rz. 226.

¹⁵⁷Vgl. Schlüter and Stolte (2016), Kap. 7 Rz. 233, 235.

¹⁵⁸Vgl. Schlüter and Stolte (2016), Kap. 7 Rz. 111.

¹⁵⁹Vgl. Schlüter and Stolte (2016), Kap. 7 Rz. 127, so auch Götz in Götz and PachHanssenheim (2016), Rz. 1075.

¹⁶⁰Vgl. Götz in Götz and PachHanssenheim (2016), Rz. 1076.

nachgehen und einen wirtschaftlichen Geschäftsbetrieb führen. Dieser beschreibt gem. §14 AO eine selbstständige und nachhaltige Tätigkeit, durch die Einnahmen generiert oder andere wirtschaftliche Vorteile gewonnen werden und die über die reine Vermögensverwaltung hinausgeht. Eine Gewinnerzielungsabsicht ist dabei nicht notwendig.¹⁶¹ Für die wirtschaftliche Tätigkeit wird der Stiftung eine steuerliche Freigrenze i. H. v. 35.000 € gewährt. Übersteigen die Einnahmen des wirtschaftlichen Geschäftsbetriebs diese Grenze, wird dieser teilweise steuerpflichtig, d. h. die Steuerbegünstigung der Stiftung bleibt im Übrigen erhalten. Voraussetzung hierfür ist allerdings, dass der wirtschaftliche Geschäftsbetrieb als Ziel die Mittelbeschaffung für den ideellen Bereich verfolgt.¹⁶² Er ist somit lediglich Mittel zum Zweck.¹⁶³ Der steuerbegünstigte Bereich der Stiftung wird also vom wirtschaftlichen Geschäftsbetrieb unterstützt.¹⁶⁴

Ein wirtschaftlicher Geschäftsbetrieb, der mit seinen Erträgen unmittelbar dem steuerbegünstigten Zweck der Stiftung dient und dessen Tätigkeit auch selbst darauf ausgerichtet ist, den steuerbegünstigten Zweck zu verfolgen, wird als steuerbefreiter Zweckbetrieb angesehen.¹⁶⁵ Voraussetzung ist, dass seine Gesamtrichtung mit geringfügigen Abweichungen von höchstens 10 Prozent dazu dient, die steuerbegünstigten satzungsmäßigen Zwecke der Stiftung zu erfüllen.¹⁶⁶ Es gelten dabei die allgemeinen Anforderungen der AO, also Selbstlosigkeit, Ausschließlichkeit und Unmittelbarkeit. Stiftungsmittel wie bspw. Spenden, Zuschüsse und Mitgliedsbeiträge dürfen zwischen dem ideellen Bereich und dem Zweckbetrieb verschoben werden. Das bedeutet, dass ein möglicher Verlust des Zweckbetriebs dadurch ausgeglichen werden kann.¹⁶⁷ Der Zweckbetrieb ist von der Besteuerung der Erträge mit Körperschaft- und Gewerbesteuer befreit. Lediglich steuerbare und steuerpflichtige Umsätze werden nach §12 Abs. 2 Nr. 8 UStG mit dem ermäßigten Steuersatz mit Umsatzsteuer besteuert.

Neben dem ideellen Bereich und dem steuerpflichtigen bzw. steuerbefreiten Geschäftsbetrieb kann die Stiftung auch innerhalb der Vermögensverwaltung aktiv werden. Die Vermögensverwaltung beinhaltet nach §14 Satz 3 AO die Nutzung von Vermögen zur Ertragserzielung, bspw. durch das Anlegen von Kapitalvermögen oder durch Vermietung/Verpachtung von unbeweglichem Vermögen. Wird das Vermögen dagegen lediglich umgeschichtet oder wird seine Substanz aufgezehrt, ist von einer wirtschaftlichen Tätigkeit auszugehen.¹⁶⁸ Die exakte Trennung dieser beiden Bereiche ist jedoch umstritten. Auf der einen Seite kann die Abgrenzung nach den einkommensteuerlichen Grundsätzen zur Trennung der Einkünfte aus Gewerbebetrieb von Einkünften aus Kapitalvermögen und aus Vermietung und Ver-

pachtung erfolgen.¹⁶⁹ Auf der anderen Seite wird diese Form der Abgrenzung nicht konsequent von der Rechtsprechung durchgehalten. Es kann auch vom Wettbewerbsschutz ausgegangen werden, der die partielle Steuerpflicht begründet und in dessen Folge die fragliche Tätigkeit zur steuerfreien Vermögensverwaltung zählt, wenn nicht mit Weitergabe des Steuervorteils an den Markt zu rechnen ist.¹⁷⁰ Eine Ausnahme zur Vermögensumschichtung stellen Käufe und Verkäufe von Vermögen in Form von Wertpapieranlagen dar. Die Analyse der Kursentwicklung und darauf gestützte Entscheidungen zur Umschichtung des Vermögens gehören üblicherweise zur Vermögensverwaltung.¹⁷¹ Die Stiftung darf allerdings nicht als gewerblicher Wertpapierhändler agieren. Dies ist abhängig von der Höhe der gehandelten Wertpapiere und der Anzahl der Veräußerungen.¹⁷² Eine unternehmensverbundene Stiftung, die als Beteiligungsträgerstiftung an einer Kapitalgesellschaft Anteile hält, betreibt dabei grundsätzlich steuerfreie Vermögensverwaltung und erhält dadurch steuerliche Vorteile. Besteht jedoch eine personelle Verflechtung zwischen den Organen der Stiftung und der der Körperschaft, ist davon auszugehen, dass die Stiftung entscheidenden Einfluss auf die Geschäftsführung des Unternehmens nehmen kann und damit bspw. die Beeinflussung des Ausschüttungsverhaltens des Unternehmens möglich ist. In diesem Fall ist von einem steuerpflichtigen wirtschaftlichen Geschäftsbetrieb auszugehen. Die personelle Verflechtung ist z. B. gegeben, wenn der Vorstand der Stiftung gleichzeitig als Geschäftsführer einer GmbH agiert, deren Gesellschaftsanteile von der Stiftung in ihrem Vermögen gehalten werden. Die Stiftung greift dagegen nicht aktiv in die Geschäftsführung ein, wenn sie als Mitglied des Aufsichtsrates des Unternehmens lediglich eine Kontrollfunktion übernimmt. Besteht die Beteiligung nicht bei einer Kapital- sondern bei einer Personengesellschaft, wird i. d. R. von einem wirtschaftlichen Geschäftsbetrieb ausgegangen.¹⁷³ Grundstücke, die zum Stiftungsvermögen gehören, kann die Stiftung vermieten oder verpachten und dabei in der Vermögensverwaltung tätig sein und keinen wirtschaftlichen Geschäftsbetrieb nach §14 Satz 3 AO ausüben, solange das Grundstück dabei nur zur Verfügung gestellt wird und keine Nebenleistungen in einem nicht unbedeutendem Umfang geleistet werden. Auch eine Veräußerung der Grundstücke ist denkbar, hierbei sollte die sog. „Drei-Objekte-Grenze“ nicht überschritten werden.¹⁷⁴ Die Vermögensverwaltung zählt wie der Zweckbetrieb zum ertragsteuerbefreiten Bereich der Stiftung, allerdings unterliegt die Vermögensverwaltung der Umsatzsteuer.

Nach §12 Abs. 2 Nr. 8 UStG findet der ermäßigte Steuersatz Anwendung, sofern keine allgemeine Umsatzsteuerbe-

¹⁶¹Vgl. Schlüter and Stolte (2016), Kap. 7 Rz. 130 f.

¹⁶²Vgl. BFH vom 23.07.2003, I R 29/02, BStBl II 2003, S. 930.

¹⁶³Vgl. Schlüter and Stolte (2016), Kap. 7 Rz. 132.

¹⁶⁴Vgl. BFH vom 04.04.2007, I R 76/05, BStBl II 2007, S. 631.

¹⁶⁵Vgl. Schlüter and Stolte (2016), Kap. 7 Rz. 139.

¹⁶⁶Vgl. BFH vom 18.01.1995, V R 139-142/92, BStBl II 1995, S. 446.

¹⁶⁷Vgl. Schlüter and Stolte (2016), Kap. 7 Rz. 140.

¹⁶⁸Vgl. Schlüter and Stolte (2016), Kap. 7 Rz. 143.

¹⁶⁹Vgl. BFH vom 27.03.2001, I R 78/99, BStBl II 2001, S. 449.

¹⁷⁰Vgl. Schlüter and Stolte (2016), Kap. 7 Rz. 143.

¹⁷¹Vgl. BFH vom 11.07.1968, IV 139/63, BStBl II 1968, S. 775.

¹⁷²Vgl. BFH vom 20.12.2000, X R 1/97, BStBl II 2001, S. 706 i. V. m. Schlüter and Stolte (2016), Kap. 7 Rz. 144.

¹⁷³Vgl. Schlüter and Stolte (2016), Kap. 7 Rz. 145.

¹⁷⁴Vgl. BFH vom 09.12.1986, VIII R 317/82, BStBl II 1988, S. 244 i. V. m. Schlüter and Stolte (2016), Kap. 7 Rz. 146 f.

freierung gilt wie bspw. bei Geld- und Kapitalverkehr.¹⁷⁵

Schwierigkeiten können sich bei der Abgrenzung von wirtschaftlichem Geschäftsbetrieb und der ideellen Tätigkeit einerseits bzw. der Vermögensverwaltung andererseits ergeben. Besonders für Tätigkeiten im Bereich des Sponsorings und bei Beteiligungen an steuerpflichtigen Kapital- und Personengesellschaften sind die von Rechtsprechung und Verwaltung entwickelten Grundsätze zu beachten. Auch die Zuordnung von gemischt veranlasseten Kosten, die sowohl den ideellen Bereich als auch den wirtschaftlichen Geschäftsbetrieb betreffen, ist nicht immer eindeutig.¹⁷⁶

Verfügt die gemeinnützige Stiftung über Grundvermögen, welches unmittelbar und ausschließlich der Umsetzung des Satzungszwecks dienlich ist, wird die Stiftung gem. §3 Abs. 1 Nr. 3b GrStG von der Grundsteuer befreit.

Im Zuge der Besteuerung bei Auflösung der Stiftung kann das Stiftungsvermögen an einen ebenfalls steuerbefreiten Rechtsträger zur Erfüllung der steuerbegünstigten Zwecke übertragen werden. In diesem Fall wird der Grundsatz der Vermögensbindung eingehalten und die Besteuerung mit Erbschaft- und Schenkungsteuer entfällt.¹⁷⁷ Wird das Vermögen jedoch an andere nicht steuerbegünstigte Personen oder Einrichtungen übertragen, löst dieser Vorgang eine Steuerpflicht aus. In diesem Fall kommt es zusätzlich zu der beim Anfallberechtigten ausgelösten Schenkungsteuer, wie bei Auflösung einer steuerpflichtigen Stiftung, zu einer rückwirkenden Besteuerung der Stiftung für einen Zeitraum von zehn Jahren, da die Gemeinnützigkeit für diesen Zeitraum aberkannt wird. Als Folge müssen im Rahmen einer umfangreichen Nachversteuerung alle gewährten Steuervorteile zurückgezahlt werden.¹⁷⁸ Wurden in diesem Zeitraum auch Spenden von Dritten eingenommen, so kommt es zur steuerlichen Spendenhaftung der Organe, die pauschal i. H. v. 30 Prozent der vereinnahmten Spenden versteuert wird.¹⁷⁹ Eine gewährte Erbschaftsteuerbefreiung i. S. d. §13 Abs. 1 Nr. 16b ErbStG entfällt, wenn die Stiftung innerhalb von 10 Jahren nach Vermögenszuwendung die Anerkennung als steuerbegünstigte Körperschaft verliert und das zugewandte Vermögen noch nicht für die steuerbegünstigten Zwecke verbraucht wurde bzw. nicht begünstigten Zwecken zugeführt wird.¹⁸⁰

4.3. Begünstigung von Zuwendungen an gemeinnützige Stiftungen

Eine Zuwendung an eine steuerbegünstigte Stiftung berechtigt den Zuwendenden zum Spendenabzug und damit mittelbar – innerhalb bestimmter Höchstgrenzen – zu einer Steuerersparnis.¹⁸¹ Für einen Abzug berechtigt sind

grundsätzlich natürliche Personen gem. §10b EStG, Personengesellschaften gem. §10b EStG und §9 Nr. 5 GewStG und Kapitalgesellschaften gem. §9 Abs. 1 Nr. 2 KStG und §9 Nr. 5 GewStG. Nach den gesetzlichen Regelungen honoriert der Staat die Förderung von gemeinnützigen Zwecken, da die Zuwendungen an gemeinnützig anerkannte Körperschaften bei der Einkommensermittlung abgezogen werden können.¹⁸² Bei natürlichen Personen handelt es sich dabei um einen Sonderausgabenabzug im Rahmen der persönlichen Einkommensteuererklärung. Für Personen- und Kapitalgesellschaften kann zum einen die Summe des gewerbsteuerlichen Gewinns um die Spende gekürzt werden, zum anderen zählt die Spende körperschaftsteuerlich zu den abziehbaren Aufwendungen.¹⁸³ Zu den begünstigten Zuwendungen zählen Spenden und Mitgliedbeiträge i. S. d. §10b Abs. 1 Satz 1 EStG, §9 Nr. 5 Satz 1 GewStG und §9 Abs. 1 Nr. 2 Satz 1 KStG. Daneben zählen auch Zustiftungen in das Stiftungsvermögen als Zuwendungen.¹⁸⁴

Zur Bewilligung des allgemeinen Spendenabzugs für Zuwendungen an gemeinnützige Körperschaften müssen bestimmte Voraussetzungen erfüllt werden. Sachliche Abzugsvoraussetzung ist die Ausstellung einer Zuwendungsbestätigung der empfangenen Körperschaft nach amtlichem Muster.¹⁸⁵ Materielle Voraussetzungen finden sich in §10b EStG bzw. bei Körperschaften in §9 Abs. 1 Nr. 2 KStG.

Als Zuwendungsempfänger kommen drei Gruppen von Empfängern in Betracht: Steuerbefreite Körperschaften; Personenvereinigungen oder Vermögensmassen; juristische Personen öffentlichen Rechts und öffentliche Dienststellen sowie ausländische Körperschaften aus einem EU/EWR-Staat.¹⁸⁶ Bei Körperschaften, die überwiegend Leistungen an ihre Mitglieder erbringen, sind die Mitgliedsbeiträge jedoch nicht abziehbar.¹⁸⁷ Auch eine in Gründung befindliche rechtsfähige Stiftung zählt nicht zu den begünstigten Zahlungsempfängern, da erst bei Anerkennung durch die Stiftungsbehörde ein Widerruf durch den Stifter ausgeschlossen werden kann.¹⁸⁸ Weitere Voraussetzung ist die Verwendung der zugewandten Mittel für steuerbegünstigte Zwecke i. S. d. §§52 bis 54 AO. Entscheidend ist hier die Zweckbindung des Zuwendenden. Wird eine Zuwendung an eine steuerbegünstigte Körperschaft getätigt, ist davon auszugehen, dass diese die Mittel für ihre steuerbegünstigten Zwecke verwendet.¹⁸⁹ Nach §10b Abs. 4 EStG ist der Spendenabzug auch bei einer widrigen Verwendung der Zuwendung möglich, soweit der Spender gutgläubig von einer Verwendung für steuerbegünstigte Zwecke ausgeht.

Damit eine Spende unentgeltlich erfolgt, darf der Zuwendende keinen besonderen Vorteil oder eine Gegenleistung erwarten, er tätigt die Spende fremdnützig und handelt um der

¹⁷⁵Vgl. Schlüter and Stolte (2016), Kap. 7 Rz. 148.

¹⁷⁶Vgl. Richter, Gollan in Meyn et al. (2013), Rz. 784

¹⁷⁷Vgl. Wigand in Wigand et al. (2015), Kap. 5 Rz. 26.

¹⁷⁸Vgl. Schlüter and Stolte (2016), Kap. 7 Rz. 118.

¹⁷⁹Vgl. Wigand in Wigand et al. (2015), Kap. 5 Rz. 27.

¹⁸⁰Vgl. Schlüter and Stolte (2016), Kap. 7 Rz. 119.

¹⁸¹Vgl. Schlüter and Stolte (2016), Kap. 7 Rz. 149.

¹⁸²Vgl. Richter, Gollan in Meyn et al. (2013), Rz. 475.

¹⁸³Vgl. Richter, Gollan in Meyn et al. (2013), Rz. 480.

¹⁸⁴Vgl. Richter, Gollan in Meyn et al. (2013), Rz. 479.

¹⁸⁵Vgl. Richter, Gollan in Meyn et al. (2013), Rz. 481.

¹⁸⁶Vgl. Schlüter and Stolte (2016), Kap. 7 Rz. 151.

¹⁸⁷Vgl. Richter, Gollan in Meyn et al. (2013), Rz. 483.

¹⁸⁸Vgl. Richter, Gollan in Meyn et al. (2013), Rz. 484.

¹⁸⁹Vgl. Schlüter and Stolte (2016), Kap. 7 Rz. 162.

Sache willen.¹⁹⁰ Keine Spenden sind bspw. Zuwendungen im Rahmen von Sponsoringaktivitäten, da hier eine Werbeleistung erbracht wird. Auch Eintrittsgelder für eine Veranstaltung, z. B. zu einem Wohltätigkeitsball, sind keine Spenden. Gleiches gilt für den Kaufpreis beim Erwerb von Gegenständen wie es z. B. bei einem Wohltätigkeitsbasar der Fall ist.¹⁹¹

Neben der Unentgeltlichkeit muss auch die Freiwilligkeit der Spende gewährleistet sein. Der Spender muss seine Zuwendung ohne rechtliche oder faktische Verpflichtung erbringen. Leistet ein Erbe eine Zuwendung an eine gemeinnützige Stiftung, zu der er aufgrund einer testamentarischen Verfügung verpflichtet ist, handelt es sich nicht um eine freiwillige Spende. Wird die Zuwendung aufgrund von moralischen oder sittlichen Gründen getätigt, kann jedoch von einer freiwilligen Spende ausgegangen werden.¹⁹²

Die Nachweispflicht zur Bestätigung, dass die Voraussetzungen für den Abzug der Spende erfüllt sind, liegt beim Spender. Dieser Pflicht kommt er nach, in dem er eine nach amtlichem Muster erstellte Zuwendungsbestätigung abgibt.¹⁹³ Ohne Vorlage einer Zuwendungsbestätigung ist ein Spendenabzug ausnahmslos nicht möglich, auch wenn die Voraussetzungen für die Ausstellung erfüllt sind.¹⁹⁴ In der Zuwendungsbestätigung muss die Höhe der Spende und der steuerbegünstigte Verwendungszweck angegeben werden.¹⁹⁵

Der Höchstbetrag für den steuerlichen Spendenabzug liegt gem. §10b EStG bei 20 Prozent des Gesamtbetrags der Einkünfte oder bei 4 Promille der Summe der gesamten Umsätze und der im Kalenderjahr aufgewendeten Löhne und Gehälter. Bei Personengesellschaften sind die Spenden bei der Gesellschaft selbst als Einkommensverwendung der Gesellschafter zu sehen. Die Spenden werden diesen anteilig zugeordnet und bei der Veranlagung berücksichtigt.¹⁹⁶

Können Zuwendungen im Veranlagungszeitraum nicht mehr berücksichtigt werden, weil der Höchstbetrag erreicht wurde, besteht die Möglichkeit, diese Zuwendungen zeitlich unbegrenzt vorzutragen.¹⁹⁷ Ein Rücktrag in das vorherige Jahr ist jedoch nicht möglich.¹⁹⁸

Eine Besonderheit erfahren Zuwendungen in den Vermögensstock einer Stiftung. Gegenüber den generellen Regelungen können natürliche Personen in den Genuss des erweiterten Spendenabzugs nach §10b Abs. 1a EStG kommen und einen Abzug i. H. v. maximal 1 Mio. € geltend machen. Dieser erweiterte Spendenabzug zählt nicht nur bei der Zuwendung an eine Stiftung im Rahmen der ersten Kapitalausstattung, sondern auch für Zustiftungen während des Bestehens

der Stiftung.¹⁹⁹ Dabei gilt für den erweiterten Spendenabzug ein Zeitraum von 10 Jahren, in dem dieser in voller Höhe geltend gemacht werden kann. Der Stifter kann entscheiden, ob der Betrag im Jahr der Zuwendung als Sonderausgabe abgezogen wird oder ob er gleichmäßig oder unterschiedlich auf das Jahr der Zuwendung und die darauffolgenden 9 Jahre verteilt wird.²⁰⁰ Der erweiterte Spendenabzug gilt zusätzlich zum allgemeinen Spendenabzug.²⁰¹ Lediglich ein Wechsel zwischen den Spendenabzügen ist nicht zulässig.²⁰² Für gemeinsam veranlagte Ehepartner gilt ein erweiterter Spendenabzug i. H. v. 2 Mio. €, wobei unerheblich ist, aus welchem Vermögen die Zuwendung letztendlich stammt.²⁰³

5. Doppelstiftung als Kombination privat- und gemeinnütziger Stiftungen

Bei dem Modell der Doppelstiftung wird eine gemeinnützige Stiftung mit einer privatnützigen Familienstiftung verbunden, um häufig eine Regelung zur Unternehmensnachfolge zu treffen. Sowohl die Familienstiftung als auch die gemeinnützige Stiftung werden als rechtsfähige Stiftungen gegründet.²⁰⁴ Dabei handelt es sich um eine Sonderform der Beteiligungsträgerstiftung.²⁰⁵ In welcher Höhe genau die Geschäftsanteile an die beiden Stiftungen übertragen werden, hängt von der Zielsetzung des Stifters ab. Wesentliches Entscheidungsmerkmal wird sein, in welchem Umfang die Familie des Stifters abgesichert sein soll. In der Regel werden alle Beteiligungen am Familienunternehmen, die nicht für die Versorgung des Stifters und seiner Familie notwendig sind, in eine gemeinnützige und damit steuerbefreite Stiftung eingebracht.²⁰⁶

Im Rahmen der Übertragung der Beteiligungen an die beiden Stiftungen wird meist festgelegt, dass die Mehrheit der Stimmrechte der Beteiligungen an die Familienstiftung übergeht. Damit wird sichergestellt, dass die unternehmerische Verantwortung noch immer beim Stifter und seiner Familie liegt und nicht auf die gemeinnützige Stiftung übertragen wird.²⁰⁷ Die Familienstiftung wird somit als sog. Führungsstiftung ausgestaltet. Auch eine disquotale Verteilung der Gewinnbezugsrechte ist steuerlich zulässig, wenn die inkongruente Verteilung zivilrechtlich wirksam geworden ist.²⁰⁸ Letztendlich werden also Vermögen und Ertrag getrennt. Hierzu muss der Gesellschaftsvertrag des Unternehmens entsprechend angepasst werden. Je nach Rechtsformwahl ergeben sich rechtsformbedingte Besonderheiten.²⁰⁹

¹⁹⁹Vgl. Schlüter and Stolte (2016), Kap. 7 Rz. 210.

²⁰⁰Vgl. Schlüter and Stolte (2016), Kap. 7 Rz. 212.

²⁰¹Vgl. Schlüter and Stolte (2016), Kap. 7 Rz. 213.

²⁰²Vgl. BMF vom 18.12.2008, IV C 4, S 2223/07/0020 BStBl I 2009, S. 17.

²⁰³Vgl. Schlüter and Stolte (2016), Kap. 7 Rz. 215.

²⁰⁴Vgl. Götz in Götz and PachHanssenheimb (2016), Rz. 1381.

²⁰⁵Vgl. Götz in Götz and PachHanssenheimb (2016), Rz. 1381.

²⁰⁶Vgl. Götz in Götz and PachHanssenheimb (2016), Rz. 1384

²⁰⁷Vgl. Wigand in Wigand et al. (2015), Kap. 11 Rz. 30, so auch G. and T. (2014), S. 130.

²⁰⁸Vgl. BMF vom 17.12.2013, IV C 2 - S 2750 a/11/10001 BStBl I 2014, S. 63.

²⁰⁹Vgl. Götz in Götz and PachHanssenheimb (2016), Rz. 1382

¹⁹⁰Vgl. BFH vom 02.08.2006, XI R 6/03, BStBl II 2007, S. 8.

¹⁹¹Vgl. Wigand in Wigand et al. (2015), Kap. 5 Rz. 123.

¹⁹²Vgl. Wigand in Wigand et al. (2015), Kap. 5 Rz. 124.

¹⁹³Vgl. Richter, Gollan in Meyn et al. (2013), Rz. 520.

¹⁹⁴Vgl. BFH vom 06.03.2003, XI R 13/02, BStBl II 2003, S. 554.

¹⁹⁵Vgl. BMF vom 07.11.2013, IV C 4, S 2223/07/0018: 005, BStBl I 2013, S. 1333.

¹⁹⁶Vgl. Schlüter and Stolte (2016), Kap. 7 Rz. 204.

¹⁹⁷Vgl. BMF vom 18.12.2008, IV C 4, S 2223/07/0020, BStBl I 2009, S.

16.

¹⁹⁸Vgl. Wigand in Wigand et al. (2015), Kap. 5 Rz. 134.

Größere Gestaltungsfreiheit bietet sich hier bei der Rechtsform der GmbH. Gem. §29 Abs. 3 GmbHG kann der Gesellschaftsvertrag vorsehen, dass die Gewinnanteile abweichend von der vermögensmäßigen Beteiligung ausgestaltet werden. Wird das Unternehmen dagegen als Personengesellschaft geführt, ist es sinnvoll, eine GmbH zwischen Unternehmen und gemeinnütziger Stiftung vorzusehen, da die unmittelbare Beteiligung der gemeinnützigen Stiftung als Mitunternehmer an der Personengesellschaft automatisch zu einem wirtschaftlichen Gewerbebetrieb führt.²¹⁰

Da die Unternehmensgewinne, die an die gemeinnützige Stiftung ausgeschüttet werden, dem Gebot der zeitnahen Mittelverwendung unterliegen, dürfen diese nicht thesauriert werden. Die Familienstiftung kann mit ihrem hohen Stimmanteil bei Beschlüssen dahingehend wirken, dass die Gewinne auf Unternehmensebene einbehalten werden, um die operativen Tätigkeiten des Unternehmens zu stärken. Das bedeutet jedoch auch, dass die Höhe der Ausschüttung an die Familienstiftung dadurch gemindert wird.²¹¹

Der wesentliche Vorteil der Doppelstiftung liegt in der Minimierung der Steuerbelastung. Da ein Großteil des Vermögens in die gemeinnützige Stiftung übertragen wird, ist diese Übertragung schenkung- und erbschaftsteuerfrei. Auch die Erträge können steuerbefreit bleiben. Außerdem ist es der gemeinnützigen Stiftung gem. §58 Nr. 6 AO erlaubt, bis zu einem Drittel ihres Einkommens für den Unterhalt des Stifters und seiner nächsten Angehörigen zu verwenden. Dies zählt i. S. d. §58 AO zu den steuerlich unschädlichen Betätigungen, die die gemeinnützigkeitsrechtlichen Begünstigungen nicht gefährden.

6. Fazit

Die unternehmensverbundene Stiftung stellt eine geeignete Gestaltung zur Lösung des Nachfolgeproblems eines Unternehmens dar. Das „Lebenswerk“ des Unternehmers und Stifters kann auf Dauer erhalten werden. Durch geeignete Stiftungskonstruktion und Regelungen in der Satzung können die Stiftungsorgane an den Willen des Stifters gebunden werden. Aber auch andere ideelle Motive können durch die Gründung einer Stiftung erfüllt werden, wie die Förderung eines gesellschaftlichen Anliegens.

Die Motive des Stifters spiegeln sich meist im Stiftungszweck wider. Es können sowohl privatnützige Zwecke wie die Versorgung der Familie oder gemeinnützige Zwecke gewählt werden. Änderungen, die nach Errichtung der Stiftung vorgenommen werden sollen, sind nicht ohne Weiteres durchführbar. Deshalb sollte bei der Konzeption der Stiftungsgestaltung das Stiftungsgeschäft inklusive der Stiftungsverfassung mit Sorgfalt erstellt werden. Dies gilt insbesondere bei Errichtung einer gemeinnützigen Stiftung.

Die gemeinnützigen Stiftungen, und eingeschränkt auch die privatnützigen Stiftungen, unterliegen der Stiftungsaufsicht durch die zuständige Landesbehörde. Es ist daneben

empfehlenswert, diese bereits bei Stiftungserrichtung zu kontaktieren und das Beratungsangebot in Anspruch zu nehmen, um im Vorfeld kritische Fragen zu besprechen.

Die Besteuerung von unternehmensverbundenen Stiftungen, die privatnützige Zwecke verfolgen, ähnelt sehr der Besteuerung anderer Körperschaften. Bei Errichtung der Stiftung kommt es meist zur Belastung mit Erb- oder Schenkungsteuer. Realisiert der Stifter stille Reserven bei der Übertragung des Vermögens, kommt es zur Besteuerung mit Einkommens- bzw. Körperschaftsteuer. Der laufende Betrieb wird je nach Einkunftsart mit Körperschaft- und Gewerbesteuer belastet. Zusätzlich kann es zur Besteuerung mit Grund- und Grunderwerbsteuer kommen. Auch Umsatzsteuer kann ggf. anfallen. Bei Auflösung der Stiftung erfolgt die Besteuerung wie bei der Errichtung in umgekehrter Weise, es kommt zur Besteuerung mit Erbschaft- oder Schenkungsteuer.

Durch eine Verfolgung von gemeinnützigen, kirchlichen oder mildtätigen Zwecken nach §52 bis 54 AO kann eine Stiftung in den Genuss von umfangreichen Steuervergünstigungen und -befreiungen kommen. Dabei wird zwischen den vier Sphären, dem ideellen Bereich, dem Zweckbetrieb, dem wirtschaftlichen Geschäftsbetrieb und der Vermögensverwaltung unterschieden. Die vom Gesetzgeber eingeführten steuerlichen Ersparnisse sollen zur selbstlosen Förderung der Allgemeinheit motivieren. Um von den Steuerersparnissen zu profitieren, müssen jedoch eine Vielzahl an Voraussetzungen bei der Mittelverwendung und andere Formalien erfüllt werden.

Als besondere Gestaltungsform kann für einen Unternehmer vor allem die Doppelstiftung in Betracht kommen. Der Stifter und seine Angehörigen erhalten Zuwendungen im Rahmen der Familienstiftung. Gleichzeitig wird eine gemeinnützige Stiftung gegründet, um von den Steuerbegünstigungen für gemeinnützige Stiftungen zu profitieren.

Natürlich muss letztendlich in jedem Einzelfall geprüft und beurteilt werden, inwiefern eine Stiftungslösung für den vermögenden Stifter oder Unternehmer vorteilhaft ist und wie die genaue Ausgestaltung dieser Lösung unter Abwägung der Eigenheiten und Erfordernisse des jeweiligen Sachverhalts aussehen kann. Bei der Beurteilung des Einzelfalls sollten auch Überlegungen zu den Gestaltungsalternativen, wie z. B. die unselbstständige Stiftung oder die Stiftungs-GmbH, Beachtung finden.

²¹⁰Vgl. Götz in Götz and PachHanssenheimb (2016), Rz. 1387

²¹¹Vgl. G. and T. (2014), S. 130.

Literatur

- H. Berndt and H. Götz. *Stiftung und unternehmen. 8. Aufl., Hamm*, 2009.
- Bundesverband deutscher Stiftungen. *Zahlen, Daten, Fakten zum deutschen Stiftungswesen*, Berlin, 2014. URL https://www.stiftungen.org/fileadmin/bvds/media/ZDF_EB00K_final_Webgr%C3%B6%C3%9Fe.pdf. Abruf am 18.04.2017.
- M. Engel. *Die unternehmensverbundene stiftung. Diss., Regensburg*, 2008.
- K. Fleschutz. *Die stiftung als nachfolgeinstrument für unternehmen - handlungsempfehlungen für die ausgestaltung und überführung. Diss., Wiesbaden*, 2008.
- Brandmüller G. and Klinger T. *Unternehmensverbundene stiftungen - recht, steuer, betriebswirtschaft. 4. Aufl., Berlin*, 2014.
- D. Gebel. *Betriebsvermögensnachfolge. 2. Aufl., München*, 2002.
- H. Götz and F. PachHanssenheimb. *Handbuch der stiftung zivilrecht, steuerrecht, rechnungslegung. 2. Aufl., Herne*, 2016.
- C. Meyn, A. Richter, C. Koss, and A. K. Gollan. *Die stiftung umfassende erläuterungen, beispiele und musterformulare für die rechtspraxis. 3. Aufl., Freiburg*, 2013.
- Ministerium für Inneres und Kommunales des Landes Nordrhein-Westfalen. *Stiftungsaufsicht*. URL <http://www.mik.nrw.de/themen-aufgaben/buergerbeteiligung-wahlen/stiftungen/stiftungsaufsicht.html>. Abruf am 03.04.2017.
- K. J. Schiffer and C. J. Schürmann. *Auflösung rechtsfähiger Stiftungen*. URL <http://www.iww.de/sb/praxisfaelle/stiftung-recht-aufloesung-rechtsfaehiger-stiftungen\discretionary{-}{-}{-}f57794>. Abruf am 28.03.2017.
- A. Schlüter and S. Stolte. *Stiftungsrecht - Formen und Errichtung, Stiftungsaufsicht, Verwaltung, Besteuerung, Internationales Steuerrecht. 3. Aufl., München*, 2016.
- Arne Schnitger. *Die Stiftung als steuerliches Gestaltungsmittel zur Sicherung des Fortbestandes eines Unternehmens?*, volume 1. Diplomica Verlag, 2006.
- Werner Seifart. *Seifart, w. and von Campenhausen, a. 2. Aufl. München*, 2009.
- Christoph Stumpf, J. Suerbaum, M. Schulte, and R. Pauli. *Stiftungsrecht: Kommentar*. München, 2011.
- Olaf Werner and Ingo Saenger. *Die Stiftung: Recht, Steuern, Wirtschaft; Stiftungsrecht*, volume 1. BWV Verlag, 2008.
- K. Wigand, C. Haase-Theobald, M. Heuel, and Stolte S. *Stiftungen in der Praxis Recht, Steuern, Beratung. 4. Aufl., Wiesbaden*, 2015.



The Effects of Battery Storage on Risk and Cost of Capital of Wind Park Investments

Lukas Erhard

Technische Universität München

Abstract

To reach the defined reduction goals for green house gas emissions, an increasing share of renewables and especially wind power is necessary. However, these generation technologies are intermittent and progressively exposed to market risks as a consequence of declining financial support in the future. To reduce revenue volatility, in this thesis, a wind farm is combined with a battery storage. The study emphasizes the battery's effect on the investment risk and the accompanying cost of capital. In order to assess this effect, I develop a deterministic optimization model based on historic wind farm and market price data in order to maximize cash flows. Monte Carlo scenarios are generated to evaluate the impact on risk by using the Value-at-Risk as risk criterion. I find that batteries can indeed reduce revenue risk in a case without subsidies. Furthermore, the link to cost of capital is made. The latter, as well as the battery prices, need to be reduced by a certain amount to make the application of a battery economically reasonable.

Keywords: Renewable energy, Energy markets, Battery storage, Wind investment, Energy investment risk

1. Introduction

The energy industry is undergoing a transformation. With negotiations on the Kyoto Protocol in 1997, climate change became more important in international politics, and leading economies around the world committed to reaching green house gas emission reduction goals.¹ The European Union (EU) has set ambitious goals that aim to reduce green house gas emissions. More precisely, by 2050 the aim is a reduction by at least 80 % compared to 1990 levels. Analyses have shown that this will only be achievable if the energy sector is close to zero carbon.² Furthermore, nuclear energy is being phased out in some countries like Germany.³ To meet these targets, the EU aims to reach a share of electricity from renewable energy sources (RES) which is at least 27 % in 2030.⁴

In order to achieve this share of RES in Europe, big investments are required. The EU Reference Scenario 2016 projects that wind power will cover 14.4 % of the total net electricity generation by 2020, whereas it projects 25 % in 2050.⁵

This underlines the importance of future wind energy investment. Expressed in numbers, the estimated required investment volume for RES from 2011 to 2030 is 1.153 billion euros.⁶

A characteristic of RES is the intermittency of electricity production. This means that the energy produced is not continuously available and depends on external factors such as weather conditions.⁷ Whereas tidal and solar power are rather predictable, accurate forecasts for wind energy are hardly feasible.⁸ Operators of conventional power plants are able to control energy production according to current market prices to maximize profit. RES however are non-dispatchable which results in obligatory sale of electricity even if the current market prices are low. To avoid this, the possibility of affordably store energy is crucial.⁹

The funding for the investments in RES comes from different investors and equity providers, who want to be paid for bearing the risk of losing money. Because of this, a certain share will be demanded as interest, namely the cost of capital. The amount of cost of capital charged will, based on the assumption of risk-aversion, depend on the accompany-

¹Cf. United Nations (2014); Oberthür and Ott (1999), p. 2.

²Cf. European Climate Foundation (2010).

³Cf. Wozabal et al. (2016), p. 688.

⁴Cf. European Commission (n.d.).

⁵Cf. European Commission (2016) p. 65.

⁶Cf. von Hirschhausen et al. (2014) p.32.

⁷Cf. Gersema and Wozabal (2016), p. 2.

⁸Cf. Hanania et al. (n.d.); Gersema and Wozabal (2016), p. 2.

⁹Cf. Gatzert and Kosub (2016) p.991.

ing risk.

There are many different risks for this kind of investments but most studies are on the same page about the most pressing risk, namely the policy design or regulatory risk. The reason for this is that the policy design and accompanying support schemes account for predictable prices and therefore reduce the revenue risk.¹⁰ The problem especially occurs to energy from RES as they are intermittent sources of energy and the electricity production cannot be controlled.

To make electricity from RES competitive, policies and accompanying support schemes have been introduced in Europe in recent years. This kind of subsidies promoted RES in the past and still continue to do so. For example, feed-in tariffs were used to provide stable revenues over the whole investment lifetime. Therefore, market risks are overturned.¹¹ Nevertheless, as RES technologies have matured and the efficiency in mastering the transformation of electricity production must be promoted, subsidies and other forms of support are intended to be cut and phased out in the future.¹²

With these curtailments of support, market risk becomes more pressing again. This leads to the necessity to cope with risk in combination with RES, especially in wind power, since the fluctuation of wind is much higher than the solar irradiation.¹³ Among other possibilities, the operation of an energy storage is a way of dealing with this risk.

Since wind energy will play an important role for reaching the RES goals, the combination of a wind power plant with an energy storage in a virtual power plant (VPP) and its ability to influence the investment risk is analyzed in this thesis. Especially dropping prices make batteries an interesting choice as energy storage.¹⁴ Most literature on this topic assesses the operation of VPPs in general from a risk-neutral and risk-averse perspective. Furthermore, the specific combination of a battery with a wind power plant is analyzed from a risk-neutral perspective only.

This thesis emphasizes the effect of a battery storage on investment risk. It will assess to which extent a battery can mitigate revenue risk coming from volatile electricity prices and stochastic generation patterns. Additionally, this impact on risk is linked to the cost of capital of wind farm investments. Consequently, the research question is the following: Can a battery storage reduce investment risk for wind farm projects, and if yes, would it be economically viable?

Section 2 of this thesis provides a review over recent relevant literature (2.1) and the definition of the research gap (2.2) this thesis is intended to fill. Subsequently, battery storage in renewable energy is introduced in chapter 3. Lead acid (3.1.1), Lithium Ion (3.1.2) and Redox-flow batteries (3.1.3) as major technologies are explained and the usage of second-life batteries for large scale energy storage is discussed (3.2).

Closing up this chapter, section 3.3 demonstrates the provision of grid services and arbitrage trading as additional strategies to generate revenues. Section 4 introduces different RES policies, especially the support schemes which are crucial for understanding the importance of this thesis' topic. Section 5 captures the topic of risks in renewable energy investments starting with an overview of different risk categories (5.1). Thereafter (5.2), the market price risk is discussed. For this reason I elaborate on the market integration of, and resulting decrease of state-aid for RES (5.2.1), the market price dependence due to direct marketing (5.2.2) and the continued operation after the support period (5.2.3). In Section 6 the theoretical link of volatility in revenues, risk and cost of capital is clarified. Section 7 represents the core of the thesis, the developed model on the operation of a wind-battery-VPP and the influence on the investment risk in a case study. It starts with an introduction to the used methodology (7.1) followed by an overview of the underlying data set (7.2) and the presentation as well as interpretation of the results (7.3). Section 8 covers alternative instruments to stabilize revenues from RES and therefore reduce investment risk. First, risk transfer (8.1) like long-term contracts with large electricity consumers are addressed (8.1.1), financial instruments in terms of energy derivatives (8.1.2) and insurance products (8.1.3). 8.2 covers possible adjustments on the market framework for RES dominated markets. 8.3 explains diversification as a way to reduce risk for RES. Finally an overview of the thesis, including the main results, implications as well as limitations is given (chapter 9). Section 2 of this thesis provides a review over recent relevant literature (2.1) and the definition of the research gap (2.2) this thesis is intended to fill. Subsequently, battery storage in renewable energy is introduced in chapter 3. Lead acid (3.1.1), Lithium Ion (3.1.2) and Redox-flow batteries (3.1.3) as major technologies are explained and the usage of second-life batteries for large scale energy storage is discussed (3.2). Closing up this chapter, section 3.3 demonstrates the provision of grid services and arbitrage trading as additional strategies to generate revenues. Section 4 introduces different RES policies, especially the support schemes which are crucial for understanding the importance of this thesis' topic. Section 5 captures the topic of risks in renewable energy investments starting with an overview of different risk categories (5.1). Thereafter (5.2), the market price risk is discussed. For this reason I elaborate on the market integration of, and resulting decrease of state-aid for RES (5.2.1), the market price dependence due to direct marketing (5.2.2) and the continued operation after the support period (5.2.3). In Section 6 the theoretical link of volatility in revenues, risk and cost of capital is clarified. Section 7 represents the core of the thesis, the developed model on the operation of a wind-battery-VPP and the influence on the investment risk in a case study. It starts with an introduction to the used methodology (7.1) followed by an overview of the underlying data set (7.2) and the presentation as well as interpretation of the results (7.3). Section 8 covers alternative instruments to stabilize revenues from RES and therefore reduce investment risk. First, risk

¹⁰Cf. Noothout et al. (2016), pp. 22, 29 f.; Cleijne and Ruijgrok (2004) p. 50.

¹¹Cf. Noothout et al. (2016), pp. 51-52.

¹²Cf. European Commission (2014), p. section 3.3.

¹³The year-on-year fluctuation of wind can be up to 25 % whereas solar irradiation is between 3 % and 5 %; Cf. Deign (2014).

¹⁴Cf. D'Aprile et al. (2016).

transfer (8.1) like long-term contracts with large electricity consumers are addressed (8.1.1), financial instruments in terms of energy derivatives (8.1.2) and insurance products (8.1.3). 8.2 covers possible adjustments on the market framework for RES dominated markets. 8.3 explains diversification as a way to reduce risk for RES. Finally an overview of the thesis, including the main results, implications as well as limitations is given (chapter 9).

2. Literature review and research gap

2.1. Literature review

Extensive literature exists focusing on risks in wind energy investments in general. Gatzert et al. for example identify the most pressing risks in wind energy investments.¹⁵ There is a further study to be highlighted, namely the Dia-Core project by Noothout et al. They analyse risks in European countries in combination with a study on cost of capital in wind energy investments in these countries.¹⁶ The main emphasis is on the relation between risk and the respective RES policies. The Green-X project by Cleijne et al. quantifies the risk in RES investments and can make the leap to the resulting weighted cost of capital (WACC) using the developed Green-X model.¹⁷

The majority of papers focuses on VPPs consisting of at least one intermittent energy source combined with another intermittent energy source or energy storage. Some of these papers e.g. Costa et al. conduct their analysis from a risk-neutral perspective and only attain the goal of maximizing profits. There also exists literature representing a risk-averse view that is based on accompanying key performance indicators for risk like the value at risk (VaR). This is due to the fact that investors will also consider risk since they are generally assumed to be risk-averse. However, the studies with risk-averse perspectives for VPPs consisting of a wind farm mainly focus on combinations with hydro power plants and consider the risk-aversion from the RES operator's perspective. Furthermore, they do not make a connection to cost of capital.¹⁸ Table 1 provides an overview of studies on virtual power plants with wind power turbines and storage. This section introduces - making no claim to be exhaustive - the most important literature as a base for the present thesis.

Pinson et al. focus on the sizing of energy storages in VPPs to hedge imbalance penalties caused by wind power forecast uncertainty. Accordingly, they assume a market, where RES suppliers are responsible for caused imbalances due to forecast uncertainties. Their model aims to determine the optimal battery size in a dynamic way at every point in time. When applied to historic data, the model provides an optimal dynamic sizing schedule dependent on the operator's risk attitude. The dynamics are included in their model to reduce

storage cost by under-utilization. Finally, they come up with the idea to introduce storage services as new independent entity in the electricity market. They consider offering storage as a service where RES producers can rent the necessary storage capacity on a daily basis.¹⁹

Gersema et al. elaborate on risk-diversification in terms of pooling different intermittent RES into a VPP. Their optimization model is based on two-stage stochastic programming. This study is not particularly emphasizing a combination of electricity storage and wind but of different RES. However, this paper has significant influence on this thesis since it takes risk preferences of risk-averse investors into account by applying the Conditional Value at Risk (CVaR). They aim to maximize utility through reducing revenue variability by selectively combining different intermittent RES. The two-stages in the stochastic optimization model represent the two decisions that must be made, and which are choosing the portfolio weights in the first stage and one year of trading decisions in the electricity market in the second stage. Gersema et al. apply the model in a case study to a market without any subsidies as well as to a market with a fixed feed-in tariff.²⁰

Liu et al., go more into detail by focusing on VPPs consisting of wind power and hydro. They investigate the coordination of such VPPs from the generating company's view in terms of bidding strategies. The study emphasizes the economic capital an electricity generating company needs to prepare to ensure continuing operation even if losses occur. It furthermore takes a risk-averse perspective by optimizing the risk adjusted return on capital (RAROC) which represents a measure that was established as an adjustment to the Return on Investment. The RAROC displays the compromise between profit and risk as it relates the expected return to the aforementioned economic capital. Five cases are carried out in total with different objectives or assumptions, namely RAROC maximization, risk-neutrality, risk-minimization, traditional risk aversion using the CVaR and the uncoordinated operation.²¹

A study comparable to Liu et al. was published by Moghaddam et al. who optimize the operation of a combined wind farm-cascade hydro system using the CVaR as risk-aversion criterion. Comparable with Gersema et al. a two-stage stochastic programming model is developed for profit maximization. The first stage decision contains the day-ahead bidding schedule. The second stage decision is related to the operation of the VPP in real time. As a result, Moghaddam et al. provide an optimal bidding strategy for the day-ahead market as well as for up- and down-regulation capacity. They find that the bidding strategy differs substantially conditional on imbalance penalty policies.²²

In their study, Ekren et al. optimize a VPP consisting of photovoltaic (PV), wind energy and a battery storage with the main objective to reduce the system's cost. The study

¹⁵Cf. Gatzert and Kosub (2016).

¹⁶Cf. Noothout et al. (2016).

¹⁷Cf. Cleijne and Ruijgrok (2004).

¹⁸There exist studies which use usual hydro plants in VPPs or pumped hydroelectric storage. The latter represents also rather a storage.

¹⁹Cf. Pinson et al. (2009)

²⁰Cf. Gersema and Wozabal (2016).

²¹Cf. Liu et al. (2015).

²²Cf. Moghaddam et al. (2013).

provides a total system view approach trying to reduce the system's total cost by obtaining the optimal sizes of the VPP elements. The results indicate how investment costs should be distributed over PV, wind turbines, battery capacity and auxiliary energy sources at the optimum.²³

Another study on wind power and battery storage is provided by Gönsch et al. who deal with the problem of time lagged commitments for RES. This study aims at providing a framework for maximizing profits under the condition of in-advance commitments. For modelling the problem of the decision when to sell and when to store energy as a Markov Decision Process, approximate dynamic programming is used. Calculations are applied for a case based on a community wind farm combined with a storage from the German supplier Bosch in northern Germany. Results show that even small battery storage with poor efficiency can increase profits substantially.²⁴

A quite comprehensive study on grid integrated energy storages was published by Doetch et al. They distinguish between storages at producer level, at consumer level and storages in the grid as well as between three scenarios. At the producer level, they consider different purposes that storages can serve, namely marketing of electricity, mitigation of feed-in management, providing balancing energy and the EEG schedule. The scenarios represent different tariffs and a post-EEG scenario that relates to wind turbines older than 20 years for which generation companies do not get any tariff or premium anymore.²⁵ To assess the different applications of energy storages, the Generic Optimization Model for Energy Storage (GOMES) with a target function to maximize profit, is developed. The returns of the VPP applying the optimized storage is then compared with the reference returns of the operation without storage. Their model shows that in the post EEG scenario and the scenario with the low tariff, the returns with storage exceed the reference returns. These results are used to determine target investment costs for energy storage that show that for the post-EEG scenario, the storage is closest to economic efficiency.²⁶

2.2. Research gap and starting point of the thesis

As shown in the previous section, extensive literature on the effect of energy storage in VPPs exists. There are existing studies with a risk-neutral as well as studies with a risk-averse perspective. Nevertheless, there is hardly any literature on the risk-averse operation of the specific combination of wind farms with batteries as energy storage. Furthermore, the studies mentioned do not include the investor's perspective by bringing together the optimized operation of the VPP with the effect on investment risk. This thesis addresses this research gap and make the leap to the cost of capital of wind farm investments. Unlike in the listed papers, no stochastic model will be developed. Instead, historical data from a

wind farm and the electricity spot market are used to build a deterministic model. The objective of this model is to optimize the operation of the VPP, analyze the influence on the investment risk and make statements on the threshold values for cost of capital to cover the additional initial investment of the energy storage.

3. Battery storage in renewable energies

To reduce risk in investments there are different instruments. A transfer of risk is often possible via long-term electricity purchasing contracts. These kinds of contracts oblige one party to supply a predetermined amount of electricity at a predetermined price and the other party to buy this amount of electricity at the agreed upon price. This way both parties can protect themselves from electricity price volatility as well as the uncertainty about prices in the future and make their business more plannable. Another way of transferring risk in RES investments very similar to those in every business is insurance. Nevertheless, there are not a lot of specific insurances for uncertainties about future cash-flows. An insurance called "lack of wind cover" is offered by the Munich Re which covers a loss of profit in case of insufficient wind.²⁷ These risk transfer instruments are emphasized in section 8. Also support policies that provide certain revenues are a form of transferring risk from the electricity producer to the society as in most cases the electricity consumers are paying for the fixed tariff or the premium via levies.²⁸ The core idea of this thesis however is mitigating or at least reducing risk via combining wind turbines with energy storage as schematically shown in figure 1. A VPP consisting of a RES and an energy storage allows to store energy produced resulting in avoidance of mandatory electricity sales at the very time the energy is produced. Therefore, this combination provides more flexibility relating the point in time the energy is sold. Assuming sufficient storage capacities, the generation company can even decouple the supply profile entirely from the production.²⁹

This section introduces and assesses different storage technologies for battery applications with RES. Subsequently, second-use batteries are emphasized and finally electricity arbitrage trading and the participation in the electricity balancing market as revenue generating strategies for grid-integrated batteries are analyzed.

3.1. Suitable battery technology

There are different types of energy storages and different ways to cluster them. One way to cluster them is by the way they store energy. Therefore, one could distinguish between mechanical, electrical, biological, electrochemical, thermal and chemical storages. An example for a mechanical storage is pumped hydroelectric energy storage, which

²³Cf. Ekren and Ekren (2009).

²⁴Cf. Gönsch and Hassler (2016).

²⁵This however will be for post 2020. Details provided in section 5.2.3.

²⁶Cf. Doetsch et al. (2011).

²⁷Cf. Gatzert and Kosub (2016), p. 990.

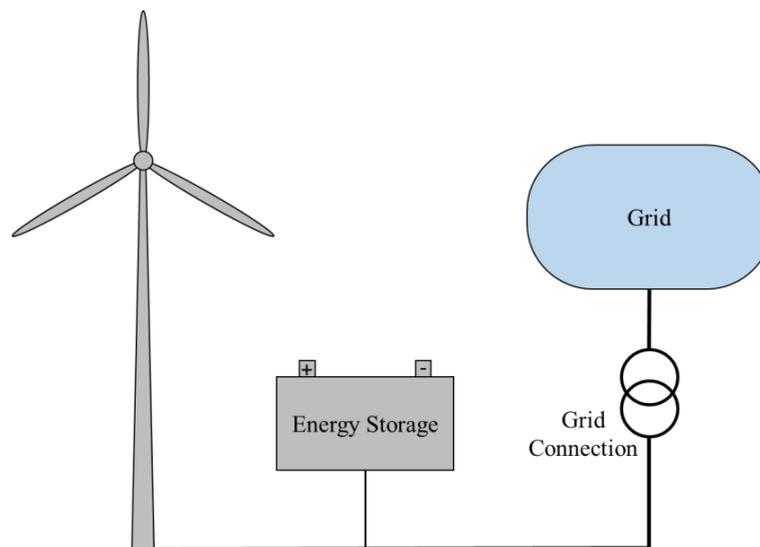
²⁸Cf. Noothout et al. (2016), p. 76.

²⁹Cf. Rugolo and Aziz (2012), p. 7159.

³⁰Own diagram based on Costa et al. (2008), p. 3.

Table 1: Literature on wind-battery virtual power plants; Source: Own table.

Author	Storage technology	Objective	Risk perspective
Pinson et al. (2009)	Not specified	Reduce imbalances, dynamic storage sizing to avoid under-utilization	risk-neutral
Gersema and Wozabal (2016)	None	Combination of RES for risk diversification	risk-averse
Liu et al. (2015)	Hydro	Optimize Hydro-Wind-VPP bidding strategy for generating companies	risk-averse
Moghaddam et al. (2013)	Hydro	Optimal profit-based risk-averse operation strategy for wind-hydro-VPP	risk-averse
Ekren and Ekren (2009)	Lead acid battery	Optimal sizing of wind-PV-battery-VPP components	risk-neutral
Gönsch and Hassler (2016)	Battery (not specified)	Profit-maximization considering time-lagged trading commitments	risk-neutral
Doetsch et al. (2011)	Battery (different technologies)	Multi-purpose grid integrated storage deployment (among others): Improve sales, provision of balancing power	risk-neutral
Costa et al. (2008)	Hydro	Increase profits and minimize imbalances of wind farms	risk-neutral
Barton et al. (2004)	Multiple, depending on strategy	Storage strategies to increase sales in limited grid connection conditions	risk-neutral

**Figure 1:** Combined wind-battery power plant³⁰

is currently the by far most widely used grid integrated energy storage and already used in VPPs with RES nowadays.³¹ Since some energy storages like e.g. pumped hydroelectric energy storage can only be realized in certain geographic conditions, electrochemical storage in the form of batteries is the chosen way to store energy used in this thesis.³² A economic operation of big battery systems is still limited by high cost,

scalability of the technologies and improvements are probably to result in a drop in costs for the future.³³

Electrochemical storages in general consist of the actual chemical storage and an electrochemical converter which converts electric to chemical energy when charging and vice versa when discharging. Batteries can be distinguished between those with external and internal storage. The most

³¹Cf. Dunn et al. (2011), p. 928.

³²Cf. Doetsch et al. (2011), p. 80.

³³Cf. Dunn et al. (2011), p. 928.

commonly used batteries nowadays are systems with internal chemical storage like lead-acid, nickel-metal hydride (NiMH), lithium-ion or sodium-sulphur.³⁴ A increasingly arising technology with external chemical storages are redox-flow batteries which consist of two reservoirs with electrolyte solutions and a membrane between electrodes past which the fluids are pumped.³⁵

In the following the main battery technologies will be introduced and respective advantages and disadvantages are addressed.

3.1.1. Lead acid batteries as a matured technology

Lead acid batteries consist of multiple cells which consist comparatively of a small number of components compared to other technologies. There is a spongy lead anode plate and cathode which is coated with lead dioxide and an insulating material used as separator in between the both electrodes. These components are placed within a plastic container which is flooded with an electrolyte consisting of a mixture of water and sulphuric acid. The separator allows the transport of the ions in the electrolyte between the electrodes and the accompanying conduction.

The biggest advantage of this technology is the high maturity level of this technology and the low-cost manufacturing. Lead acid battery capacity is the cheapest compared to other rechargeable cells. The technology is quite robust and tolerant to overcharging.

When it comes to disadvantages, the limited timespan in matters of cycles and the fact that these types of cells are not suitable for fast charging must be mentioned. Furthermore, there are some environmental issues with respect to the used materials like lead and acid.³⁶

3.1.2. Characteristics of lithium ion batteries

Within a Lithium Ion battery the electrodes form host structures where the anode has a graphene structure and the cathode is a lithium-intercalation compound in a layer structure. By inserting or removing lithium ions in or from the electrodes' structure, the battery can be charged or discharged. When charging, lithium ions are removed from the layered intercalation compound and intercalated into the graphene structure and vice versa when discharging. An electrolyte which transports the lithium ions and separates the anode from the cathode is located in between the electrodes.³⁷

A big advantage and the reason for its use in mobile applications is the high energy density and thus corresponding high specific energy and power.³⁸ Furthermore lithium ion batteries require relatively low maintenance and suffer significantly less from self-discharge compared to some other technologies. As a disadvantage the manufacturing of the

cells still is expensive and the battery requires protection to ensure operation within safe limits as the technology is not as robust as comparable rechargeable technologies.³⁹ However, strongly decreasing prices are forecasted for this technology in the future.⁴⁰

3.1.3. Introduction to redox-flow batteries

Flow batteries differ from other battery technologies as they do not have internal solid electrodes as energy storage. A redox-flow battery uses two electrolyte solutions which are stored in external tanks and pumped through the cell continuously. Inside the cell, an ion-selective membrane separates the two compartments where the two accordingly charged or discharged electrolyte solutions flow through. While redox-active ions undergo reduction or oxidation reactions when close to the electrode, non-reaction ions can still pass the membrane to ensure electrolyte balance.⁴¹

One big advantage of this technology is the independence between storage and converter when it comes to sizing. This is a difference to systems with an internal storage.⁴² The modular structure allows smooth extensions of the storage capacity and is not limited in discharge depth like other technologies are. Furthermore, there is very little maintenance required.⁴³ Another advantage is the relatively simple electrode reaction compared to other batteries. Mentionable disadvantages are the necessity of pumps, reservoirs and sensors as well as the lack of technological maturity in comparison to other technologies.⁴⁴

3.2. Potential of second-use electric vehicle batteries

Since the share of electric vehicles (EV) is increasing, multiple studies introduced the idea of using their integrated battery storages when connected to the grid.⁴⁵ Although using this service called vehicle-to-grid, for a VPP has advantages like a variable storage capacity and the mitigation of large initial investments, in this thesis stationary storing will be assessed due to simpler determination of storage cost.

Batteries suffer from slight capacity losses over time which is a major issue for application in EVs, since a high energy density is necessary. Nevertheless, these batteries are still suitable for stationary power storages as the mentioned capacity losses are not of major importance for this usage. A joint project of Daimler, the Mobility House, GETEC and REMONDIS shows the potential of second-use batteries. For this project 1,000 battery systems from used cars were retrofitted and combined to the worlds' largest second-use battery storage with a total capacity of 13 MWh. The task of this storage is to provide primary controlling power and the

³⁹Cf. Poole (2017b).

⁴⁰Cf. D'Aprile et al. (2016).

⁴¹Cf. Dunn et al. (2011), p. 933.

⁴²Cf. Doetsch et al. (2011), p. 64.

⁴³Cf. Energiespeicher - Forschungsinitiative der Bundesregierung (2015).

⁴⁴Cf. Dunn et al. (2011), p. 933.

⁴⁵Cf. e.g. Kempton and Tomić (2005); Lassila et al. (2012); Vasirani et al. (2013).

³⁴Cf. Doetsch et al. (2011), p. 64.

³⁵Cf. Dunn et al. (2011), p. 933.

³⁶Cf. Poole (2017a).

³⁷Cf. Dunn et al. (2011) p. 930.

³⁸Cf. Dunn et al. (2011) p. 930.

batteries are estimated to be cost-efficient in this operation for at least ten years.⁴⁶

3.3. Arbitrage trading and provision of grid services with battery storages

Energy arbitrage trading represents the strategy to make use of the spread between electricity prices at different times. Revenues are generated by charging the battery with bought energy at low prices and selling, which consequently equals discharging the storage in periods of high prices.⁴⁷ The time horizon for electricity arbitrage trading however is restricted due to technological characteristics and capacity. This means that high capacity storages like a large-scale pumped hydro may benefit from more long-term spreads such as the weekly load pattern. This spread is affected by the difference in demand between working days and weekends. However, batteries discharge over shorter time and are more applicable to time horizons in arbitrage trading within one single day. This implies revenue generation by taking advantage of the differences between on-peak and off-peak periods during the day.⁴⁸ The increasing market penetration of RES influences the electricity price spread. Whereas PV produces electricity during the day-time and consequently is lowering the peak prices, the wind power feed-in pattern is less systematic and is increasing the spread.⁴⁹ Nevertheless, the revenues generated by electricity arbitrage trading nowadays are too low to cover the cost of an appropriate battery storage.⁵⁰

Furthermore, it is conceivable to generate revenues by providing services to the grid using a battery storage. Services that can be provided by energy storages include e.g. the provision of electric supply capacity, reserves or frequency regulation. Providing electric supply capacity aims at reducing system peaks by shifting electricity depending on demand. This can possibly replace specific plants which are solely operated to ensure power supply during peak hours. The supply of reserves covers generation and demand deviations to ensure grid stability.⁵¹ This also includes fast-responsive capacities which must be available within ten minutes to compensate outages of generation or transmission facilities. Reserve that is available to sustain the system's frequency needs to respond within ten seconds.⁵² By providing the listed services, revenue streams can be generated.⁵³

These revenue generating strategies are just emphasized to point out that there are additional ways of operating grid integrated energy storages. However, the focus in this theses is the deployment in terms of a VPP to stabilize revenues and reduce risk.

4. Renewable energy policy and support schemes

As mentioned in the introduction, to reach defined green house gas emission goals, the share of RES in the energy market certainly has to increase significantly. Since the market mechanisms itself would fail to deliver the required share of RES, authorities need to intervene.⁵⁴ In Europe, the respective countries specify the applied support schemes with guidance by the European Commission.⁵⁵

The principles of support schemes are generally distinguished between investment-based and generation-based support. For investment-based support schemes, investments are subsidized in the form of e.g. tax reductions or soft loans. Generation-based support instruments have the task to reduce price and quantity risk of RES.⁵⁶ The latter instruments are crucial for the market-integration of RES since they address the market risk. Consequently, it is not further elaborated on investment-focused support. Figure 2 provides an overview of the different support schemes and ranks them in order to the accompanying price risk. Additionally to the different support schemes, this section briefly covers the topic of balancing responsibility and curtailment.

The first group of support schemes shown in figure 2 are the feed-in tariffs (FIT). For the fixed price FIT, RES operators receive a fixed tariff that is independent of the market price movement. This leads to the lowest, theoretically non-existent price risk. This ensures predictable, stable revenues since a constant tariff is usually provided over the project's predefined lifetime.⁵⁸ In some countries the tariff FIT is not completely fixed but varying with season or time of the day. However, weather dependent RES like wind are exempted from these constraints.⁵⁹

The second group are premiums. The latter are more market-oriented compared to FIT since power producers participate in the market and receive a premium paid on top of the wholesale price. The allowance of market signals is a substantial advantage of premium systems since it gives incentives to feed in energy when prices and thus demand is high.⁶⁰ The height of the premiums can be fixed or sliding. However, the latter has the lower price risk since it smoothens the total remuneration compared to the volatility of the market prices whereas fixed premiums pass on the price profile.⁶¹ For sliding feed-in premiums, the RES producer gets the difference between the spot price (for which the electricity is marketed) and a guaranteed price. This implies that the premium is zero when the market price equals or exceeds the strike price. Therefore, the price risk of a sliding feed-in premium is comparably low. Nevertheless, the electricity producer has to participate in the market. As a consequence,

⁴⁶Cf. The Mobility House (2016).

⁴⁷Cf. Salles et al. (2016), p. 1.

⁴⁸Cf. Staffell and Rustomji (2016), pp. 213, 215.

⁴⁹Cf. Staffell and Rustomji (2016), p. 2013.

⁵⁰Cf. Staffell and Rustomji (2016), p. 224.

⁵¹Cf. Günter and Marinopoulos (2016), p. 227.

⁵²Cf. Eyer and Corey (2010), p. 31.

⁵³Cf. Günter and Marinopoulos (2016), p. 234.

⁵⁴Cf. European Commission (2013), p. 3.

⁵⁵Cf. European Commission (2017).

⁵⁶Cf. Binda Zane et al. (2012), p. 79.

⁵⁷Figure from Binda Zane et al. (2012), p. 79.

⁵⁸Cf. Noothout et al. (2016), p. 51.

⁵⁹Cf. Binda Zane et al. (2012), p. 81.

⁶⁰Cf. Klessmann et al. (2008), p. 3656.

⁶¹Cf. Binda Zane et al. (2012), p. 81.

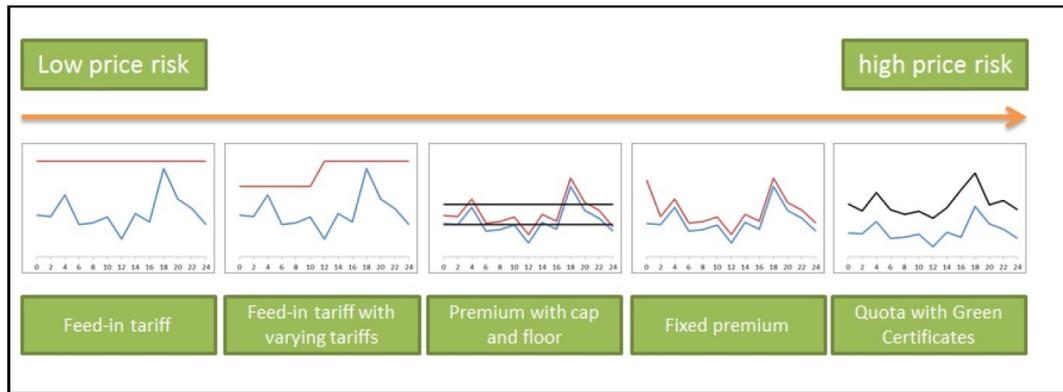


Figure 2: Support schemes with classification of the accompanying price risk⁵⁷

other requirements like forecasting the feed-in quantity have to be met.⁶²

Further forms of sliding premiums exist where e.g. no premium is paid when negative electricity market prices occur. This transmits the explicit market signal of supply surplus.⁶³ Another type of premium is the cap-and-floor premium shown in figure 2. This scheme has an increased price risk compared to regular sliding feed-in premiums since premiums are fixed for market prices within a certain range, limited by the cap and the floor. For the case that market prices rise to a level where the total remuneration, consisting of the market price plus the premium, would exceed the cap, the premium is decreased to limit the remuneration to the price cap. If the market price itself goes beyond this cap price, no premium at all is paid. However, the pre-defined floor price guarantees a certain minimum revenue.⁶⁴ This support strategy limits the price risk on the one hand but also profits on the other hand.⁶⁵

Fixed premiums are independent from the market price and bear the highest price risk among the feed-in premium support schemes. They transmit fluctuations of market prices directly to revenues. The consequences are uncertain and unstable cash flows levels allowing for high profits or losses.⁶⁶

Next to feed-in support schemes, quota regulations exist. The latter result from the definition of a certain minimum quota of RES an electricity supplier must have in its energy mix. These quotas can be technology specific to stimulate the expansion of particular technologies. In addition to the market price, RES producers receive renewable energy certificates for sold units which can be traded on a specific market. This leads to a total remuneration dependent to both, price fluctuations from the electricity market and the green certificate market. The comparatively high price risk results from the dependence on two independent market price risks.⁶⁷

Some countries like Germany have a priority feed-in policy for electricity from RES. This leads to the fact that in a competitive market, renewable energy is preferred over electricity from other sources. As a result, producers of electricity from RES can be certain that their energy is bought.⁶⁸ Nonetheless, there are plans of the German government of partially eliminating the priority feed-in policy in Germany to level the playing field for all energy sources.⁶⁹ Other countries like Estonia or Finland do not even have this policy. Without priority dispatch, operators face uncertainty of finding a counterparty in the competitive market that buys the produced electricity. The risk of not being able to sell the output and be curtailed instead is called volume risk.⁷⁰

Policies also specify the responsibilities for participating in the market. The two key types of energy markets are the spot market and the balancing market. Usually 12 - 48 hours in advance, supply and demand bids must be submitted in the spot market and by finding the equilibrium, the prices are determined for a whole day consisting of 24 hours. For the case that the forecasts, which the supply bids are based on are not met, the deviation between supply and demand must be regulated at the balancing market.⁷¹ Since especially wind is intermittent, the balancing risk is another substantial risk for RES operators. This arises due to the obligation forecasting the amount of electricity that will be fed into the grid which equals the planned production. For deviations from the forecast, penalties apply or energy has to be bought from the balancing market. Plant operators often pass on the balancing responsibility to transmission system operators which charge a specific margin for underwriting this risk. However, in several countries, intermittent RES are exempt from the balancing responsibility.⁷²

However, countries handle support schemes and exemptions differently. Sometimes, specific support schemes are only open to generators up to a certain size and support pe-

⁶²Cf. Noothout et al. (2016), p. 52.

⁶³Cf. Noothout et al. (2016), p. 51.

⁶⁴Cf. Binda Zane et al. (2012), p. 82.

⁶⁵Cf. Klessmann et al. (2008), p. 3656.

⁶⁶Cf. Noothout et al. (2016), p. 52.

⁶⁷Cf. Binda Zane et al. (2012), pp. 82-83; Klessmann et al. (2008), p. 3565.

⁶⁸Cf. Binda Zane et al. (2012), p. 84.

⁶⁹Cf. Becker (2016).

⁷⁰Cf. Binda Zane et al. (2012), p. 84.

⁷¹Cf. Krohn et al. (2009), p. 93.

⁷²Cf. Binda Zane et al. (2012), pp. 84-86.

riods can vary. Furthermore, countries exist, where generators can choose between different subsidy forms. The support schemes usually evolve and are adjusted after a while.⁷³

The mentioned policies for RES reduce risks coming from market prices, volume or balancing substantially. However, RES technologies have matured and are getting more grid-competitive. To reach the aimed renewables' share in a cost-effective way, the mentioned exemptions and financial support schemes are planned to be phased out.⁷⁴ As a consequence, RES generators will be increasingly exposed to market risks in the future.

5. Risks in renewable energy investments

The following chapter firstly emphasizes risks in RES investments in general and outlines why exposure to the electricity market is crucial. The second part demonstrates circumstances which lead to advanced market volatility dependence.

5.1. Risk categories and ranking

In literature, there are some studies on risks in RES investments which cluster possible risks in categories as shown in table 1. The classifications and definitions of the categories are slightly different. Nevertheless, policy and regulatory risks seem to be among the most pressing concerns in these studies.

Policy design risk is very important as the height and the calculability of the projects' returns strongly depend on the support schemes provided from the government. Investors of RES projects rely on specific support schemes and policies like those introduced in the section above. The financial support usually is guaranteed over a predetermined timeframe which can e.g. be 20 years. The returns of an investment often are planned based on promised subsidies.

The situation in Spain showed the importance of this risk. Policies and support schemes were retrospectively changed. Consequently, the RES investments did not have the expected returns with which they calculated and which were considered as safe. These policy actions resulted in distrust about the politics which lead to defaulted wind energy funds. Furthermore, a period without any new investments in RES and a high cost of capital for investments occurred.⁷⁵

A case like this is quite unlikely in countries like Germany, where the policy support schemes are anchored in law like the German Renewable Energy Sources Act (EEG). This beneficially influences the cost of capital which is consequently comparably low.⁷⁶

The crucial about regulatory risk compared to e.g. technological risk is that there are hardly any internal measures to mitigate this risk. Only policy makers influence it.⁷⁷

Changes in governments' priorities and accompanying changes in budgets can result in modifications of RES support. As these priorities are a political issue and may change over time, there is uncertainty about prospective support as well as retrospective adjustments like happened in Spain.⁷⁸

The nature of policy design risk will probably change in the future as RES are expected to reach grid-competitiveness by 2030. To get a transition to a cost-effective energy delivery, RES must compete in the market and support schemes like feed-in tariffs or exemptions from balancing responsibilities have to be phased out.⁷⁹ With absence of support, intermittent RES have to be marketed directly, which leads to penalties for imbalances and deviations from the prediction. Furthermore, variable revenues occur due to price volatility. These changes of state-aid are of special relevance if no or not sufficient energy storage is available. This may lead to mandatory energy sales at low prices.⁸⁰

Consequently, an energy storage cannot reduce the policy or regulatory risk itself. But these risks however are clearly related to the market or sales risk, which RES are more exposed to without or with instable support schemes. This means that a wind farm whose support scheme is abolished retrospectively thereafter is exposed to volatile market prices. Nevertheless, the market price volatility is not the only variability factor of the revenues of a wind farm. As the revenue consists of both, the quantity of electricity sold and the market price at this point in time, the wind speed fluctuations and the accompanying intermittency of output represent another variability in the revenues. Aside from short-term wind speed fluctuations there are also variabilities in the overall annual wind. The deviations of a single year wind yield from the long-term average can be substantial. Nevertheless, debt lenders and investors still demand their interests or returns, even in poor years regarding wind yield.⁸¹ The mentioned price and output volatilities however, are not independent from each other but related. Since the share of wind energy increases, there is a strong supply surplus in windy areas or at windy times, resulting in depressed market prices.⁸² There appears to be a tendency to negative correlation between market prices and available wind speeds which means low prices if there is a lot of electricity from wind to be marketed and vice versa.⁸³ Concluding it can be stated that the price variance is increasing for an increasing amount of RES installed.⁸⁴

However, aside from the risk of increasing exposure to the market and output fluctuations due to policy changes, there are already nowadays application cases for storage capabilities. These cases are e.g. direct marketing under feed-in premiums, where higher revenues might be generated using a storage. Another case is post-support operation, which

⁷⁸Cf. Gatzert and Kosub (2016), p. 995.

⁷⁹Cf. European Commission (2014), p. section 3.3.1.

⁸⁰Cf. Gatzert and Kosub (2016), p. 984.

⁸¹Cf. Rathmann et al. (2011), p. 75.

⁸²Cf. Hirth (2015), p. 156.

⁸³Cf. Sioshansi (2011), pp. 1-2.

⁸⁴Cf. Wozabal et al. (2016), p. 705.

⁷³Cf. Krohn et al. (2009), pp. 81-86.

⁷⁴Cf. European Commission (2014), section 3.3.

⁷⁵Cf. Noothout et al. (2016), p. 148 f.

⁷⁶Cf. Noothout et al. (2016), p. 3.

⁷⁷Cf. Gatzert and Kosub (2016), p. 996.

Table 2: Risk categorization; Source: Own table based on Noothout et al. (2016), p. 21; Gatzert and Kosub (2016), p. 984; Cleijne and Ruijgrok (2004), p. 15.

DiaCore	Gatzert	Green-X
Social acceptance risk	Strategic/business risk	Operational risk
Administrative risk	Transport/construction/completion	Product market risk
Financing Risk	Operation/maintenance	Input risk
Technical \& management risk	Liability/legal risk	Regulatory risk
Grid access risk	Market/sales risks	Financial risk
Policy design risk	Counterparty risk	
Market design \& regulatory risk	Political, policy, regulatory risks	
Sudden policy change risk		

means continued operation of wind turbines after the period of state aid.

5.2. Dependence on electricity market prices as a major risk

Being dependent on market prices implies variability in returns and consequently represents a pressing risk for both, RES operators and investors. The following section shows in which ways RES are increasingly exposed to market price fluctuations as a result of the targeted market integration of RES. Additionally, it is discussed, in which way an energy storage can offer a certain additional value to a wind farm.

5.2.1. Declining subsidies for renewable energy

To make RES more cost-effective, the European Commission wants subsidies in any form to be phased out in a degressive way in the future. Renewable energy is wanted to be integrated in the energy market as these sources of energy will be grid-competitive in the future.⁸⁵ To represent the maturing and learning of the technologies and to avoid overcompensation, there are predefined degression steps for the paid premiums or tariffs.⁸⁶ Furthermore, competitive bidding process as market instrument for market-premiums in direct marketing is introduced in Germany in the latest Act on the Development of Renewable Energy Sources for 2017. This means that only the RES operator with the lowest bid will get the market-premium.⁸⁷ Besides the goal of more cost-effective RES through more competition, competitive bidding helps for better plannability of capacity added. This is becoming more important as there are already sometimes bottlenecks in the power grid.⁸⁸ This application of tenders is one further step of phasing out the support provided by the government in form of feed-in premiums.

In the course of the market integration, there are also plans to cut the privilege of priority feed-in for RES.⁸⁹ A

combination of a battery with a wind farm could be the technological advantage to operate a wind farm more efficiently compared to standalone plants and therefore enable operation with less financial support. Consequently, this can lead to being awarded for the state-aid under a tender scheme with competitive bidding.

The almost grid competitive RES will probably not get any subsidies in the more distant future. Additionally, operators will have to deal with trading their energy at the electricity market. When being completely exposed to the market, an energy storage likely can improve revenues of a plant by shifting the feeding in of produced electricity from a low-price period to a time with higher prices. The added value of a storage increases with more volatile market prices. This is due to the fact that the additional revenues represent the price spread multiplied with the given quantity and subtracted the losses due to storing.

5.2.2. Dependence on price volatility through direct marketing

One way of selling produced energy is e.g. the direct marketing in Germany according the section 34 of the RES Act.⁹⁰ Other than for FIT, direct marketing means that the energy producer sells energy directly and gets paid a market premium. The height of the market premium equals the difference of the average electricity price and a fixed feed-in tariff which is granted to older and small plants. The electricity price average is usually multiplied with a different profile factor for different RES.⁹¹ Consequently, RES operators marketing their energy directly are supposed to have the same height of revenues in average as with fixed feed-in tariffs. Assuming, the operator would be able to store energy produced in low electricity price periods and sell it at a later point in time for a higher price, its average price at which sold its electricity is sold is higher than the markets' average price in the same month. Since the operator receives the

⁸⁵Cf. European Commission (2014), p. section 3.3.1.

⁸⁶Cf. Held et al. (2014), p. 5.

⁸⁷Cf. Bundesministerium für Wirtschaft und Energie (2016), p. 6.

⁸⁸Cf. Bundesministerium für Wirtschaft und Energie (2016), pp. 13, 14.

⁸⁹Cf. Becker (2016).

⁹⁰RES Act in Germany is the EEG.

⁹¹Cf. Nicola (2017a).

same market premium as other operators who do not have the possibility to store energy, higher revenues will be generated. Using this output controlling strategy, revenues can even exceed the fixed feed-in tariff level. Figure 3 illustrates, how a RES operator can increase its total remuneration at a premium that is fitted to the overall average market prices by increasing his average market price at which the electricity was sold. Of course, this approach is limited in some matter as the month' average electricity price which is used to determine the market premium will change if more RES operator use a storage strategy.

As mentioned in section 3, for receiving feed-in premiums, operators need to market their output. This usually is accompanied by balancing responsibility as every market participant must forecast the quantity feeding in and balance arising deviations. RES operators can transfer the balancing risk to direct marketers. This is done regularly as self-balancing is challenging for intermittent RES. The direct marketers assume the balancing responsibility e.g. by marketing electricity for a network consisting of multiple plants which leads to lower balancing needs. This third party, namely the direct marketers usually ask for a margin about 0.2 euro cents per kWh.⁹³ For the time between 2013 and 2015 the overall average electricity price at the day-ahead spot market was around 34 € .⁹⁴ Consequently, about 6 % of the electricity price in average represent the margin that has to be paid for marketing.⁹⁵ By using an energy storage, a wind farm is likely to be able to self-balance the output and therefore save some of these expenses.

5.2.3. Post support operation

Another scenario in which wind turbine operators are dependent on the market mechanisms is the post support operation. This describes the operation of a wind park after the period in which incentives or a sort of subsidies from the government are granted. As an example, the guaranteed duration of governmental support in Germany is 20 years.⁹⁶ Since the extension of wind energy took place in the nineties, the topic of post support operation will gain in importance in the upcoming years. In 2016, more than 7,000 wind power turbines in Germany reached an age between 15 and 20 years. This number will even increase and in 2019 there will be more than 10,000 reaching this age.⁹⁷ Due to the fact that the first EEG in Germany came into force in 2000 and all wind power turbines that already operated before this year were also guaranteed 20 years of state-aid from the year 2000 on, the financial support will expire for a significant number of plants in the end of 2020. This number is assumed to lie between 5,600 and 7,000 wind power turbines in Germany.

Furthermore, there will be about 1,600 wind power turbines with support period expiring at each of the following years.⁹⁸

For Germany, wind power plants are designed and planned to operate for 20 years. Although, plants which started operation before 2000 will get the promotion further on, there has to be a report which certifies that continuing operation is possible without any concern. Apart from this certification, there are further requirements to meet which result in additional cost.⁹⁹ As investments like RES are normally depreciated over a long span of time e.g. 15 years, older wind turbines can often draw from their full economic potential. Having a wind farm which is completely depreciated makes a continued operation even more attractive.¹⁰⁰ Fully depreciated plants can generate cheap electricity. Nonetheless, arising expenses for operation including lease, insurances, maintenance must be covered by the revenues. If this is the case, depends on the evolution of the electricity market prices in the first hand. For continuously low prices and imperfect circumstances of a project, an economically feasible operation after 2020 is probably not possible.¹⁰¹

Considering the deployment of an energy storage in combination with a post-support wind farm to shift produced energy to periods with sufficiently high prices can possibly make the continued operation reasonable.

6. Revenue volatility and its influence on cost of capital in theory

Projects in wind energy have a characteristic cash flow course which is displayed in figure 4. At the beginning of a project, the initial investment such as development expenditures and construction cost is required. During the operation, energy sales minus operational cost, interest, tax and the repayment rates represent the free cash flow. Finally, decommissioning expenses arise at the lifetime's end. This curve provides the baseline for a project's valuation.

Valuing projects as basis of decision-making is often based on the project's Net Present Value (NPV). This performance indicator is based on the concept of time value of money. The latter states that a certain amount of money today is worth more than the same amount in the future. This is due to losing the opportunity of earning interest for alternatively investing this money during the project's period.¹⁰³ The NPV can be calculated by using the Discounted Cash Flow (DCF) method. For this method, future cash flows are projected and discounted with a discount rate and offset with the initial investment of the project.¹⁰⁴ The difference between the

⁹²Own figure.

⁹³Cf. Wallasch et al. (2015), p. 19.

⁹⁴Based on EPEX Spot Day-Ahead prices for 2013 – 2015.

⁹⁵The real cost of marketing depend also on factors like park size, location, contract term (Cf. Wallasch et al. (2015), p. 19).

⁹⁶Cf. Nicola (2017b).

⁹⁷Cf. Wind-Turbine (2016a).

⁹⁸Cf. Wallasch et al. (2015), pp. 1-2.

⁹⁹Cf. Wallasch et al. (2015), pp. 6-9.

¹⁰⁰Cf. Wind-Turbine (2016b).

¹⁰¹Cf. Wallasch et al. (2015), pp. 30-32.

¹⁰²Figure from Noothout et al. (2016), p. 59; DEVEX stands for Development expenditures and equals the development cost; CAPEX, the capital expenditure, is the initial investment; OPEX are operational expenses that arise in relation to the operation.

¹⁰³Cf. Welch (2009), p. 18.

¹⁰⁴Cf. Wall Street Oasis (n.d.).

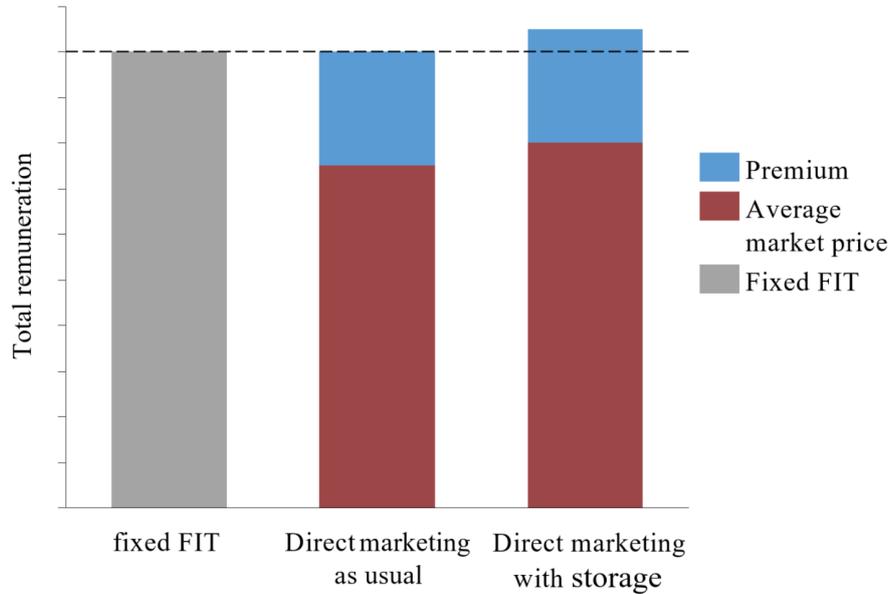


Figure 3: Comparison of total remuneration with and without storage⁹²

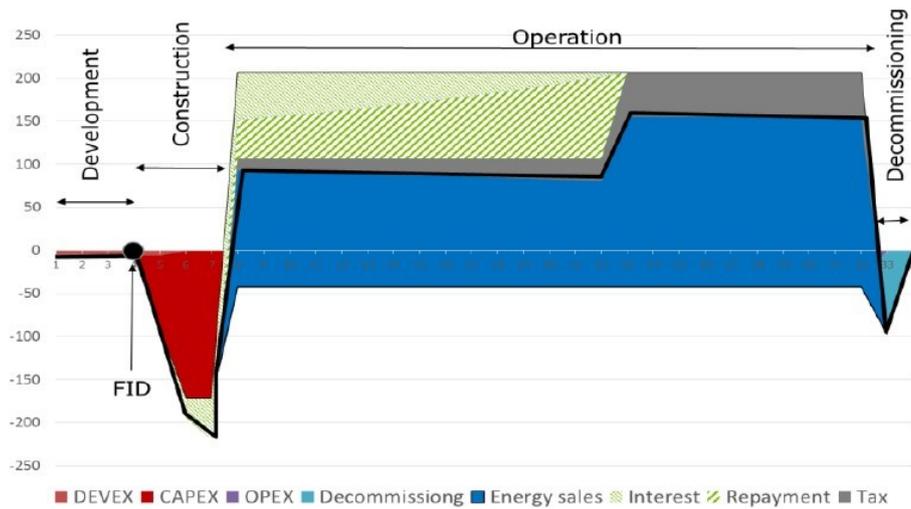


Figure 4: Characteristic cash flow of a wind energy project¹⁰²

Present Value (PV) and the NPV only lies in the consideration of the initial investment for the NPV which usually is a negative upfront cash flow.¹⁰⁵ The NPV is calculated the following way¹⁰⁶

$$NPV = C_0 + \sum_{t=1}^T \frac{C_t}{(1+r)^t} \quad (1)$$

C_0 represents the initial investment at $t = 0$, and is for this reason usually a negative figure. The sum of the discounted cash flows for the respective periods from $t = 1$ to the project's last period T is added to C_0 . $(1+r)^t$ is the applied discount factor and R is the discount rate which reflects

the project's cost of capital. This implies that a project with lower cost of capital can result in a higher NPV than another project with the same cash flow projections. The commonly used method to determine the cost of capital is the WACC. The latter weights the cost of equity and cost of debt with a project's financing structure. A project often is financed both, by equity from investors and debt from banks or lenders. The following formula shows the calculation of the WACC¹⁰⁷

$$WACC = (r_d * 1 - T) * \frac{D}{D + E} * r_e * \frac{E}{E + D} \quad (2)$$

¹⁰⁷Based on Rosenbaum and Pearl (2013), p. 151.

¹⁰⁵Cf. Welch (2009), p. 35.

¹⁰⁶Based on Welch (2009), p. 35.

where r_d = cost of debt
 r_e = cost of equity
 T = marginal tax rate
 D = marketvalue of debt
 E = marketvalue of equity

Expressed in words, the WACC is adding the components of cost of equity for the equity share and the cost of debt for the debt share. Therefore, $\frac{D}{D+E}$ represents the share of debt and $\frac{E}{D+E}$ the share of equity in the capital structure. These are the weights for averaging the cost of capital. Furthermore, the interest payment for debt r_d is considered as expense and consequently tax deductible which has a lowering effect on the WACC.¹⁰⁸ If a project is 100 % equity financed, the WACC consequently equals the cost of equity and vice versa for debt financed projects.

When calculating the WACC, the project's marginal tax rate as well as the capital structure are usually given. For new projects, an optimal target capital structure can be determined.¹⁰⁹ A study that analyzed debt-to-equity ratios across Europe found that countries with comparably low-risk environments allow for higher debt ratios. According to this study the ratios reach from 50/50 in eastern European countries like Romania and Bulgaria up to 80/20 in countries like France, Germany or Denmark.¹¹⁰ Generally speaking debt is the cheaper form of capital compared to equity.¹¹¹ This is due to the fact that debt is paid off first in the case of financial distress when assets have to be liquidated.¹¹² The tax deductibility of cost of debt makes debt even more attractive. Therefore, a higher share of debt usually decreases the WACC. In theory, the WACC decreases with an increasing proportion of debt up to a threshold value where the optimal capital structure is reached. Beyond this critical ratio the debt share is so high that financial distress becomes more likely. Consequently, debt lenders as well as equity providers will likely demand for higher interest since risk increases.¹¹³ Therefore, the risk assessment of projects determines which debt ratios are allowed at which interest rates. Knowing the fact that debt has lower cost compared to equity, high debt ratios like in Germany or Denmark are highly beneficial for financing RES projects.¹¹⁴

With a defined marginal tax rate and debt-to-equity ratio the cost of debt and the cost of equity must be determined in the next step. Since the cost of equity represents the investors risk perception, it can be used as investment risk proxy.¹¹⁵ The height of these cost can vary across different investors even within the same project. Furthermore, it is implicit and therefore not observable like the interest rate for debt.¹¹⁶ To

determine the cost of equity, the Capital Asset Pricing Model (CAPM) is often used. For the latter, the cost of equity equals the expected return from the investors perspective. The calculation of the cost of equity r_e according the CAPM can be seen in (3)¹¹⁷.

$$r_e = r_f + \beta(r_m - r_f) \quad (3)$$

where r_f = risk-free interest rate
 β = Beta
 r_m = expected market return
 $(r_m - r_f)$ = market risk premium

The risk-free rate represents the expected payoff from an investment which is considered as risk-less. Usually defined as risk-free are investments where neither default risk nor uncertainty about reinvestment rates exist. An usually applied proxy for the risk-free rate are U.S. government bonds.¹¹⁸ The market risk premium corresponds to the spread between the expected return on a market portfolio and the risk-free rate. Since investors usually are risk averse, higher returns are expected when investing in the market compared to the risk-free rate. The expected market return can be deduced from surveys, historical data or current market data.¹¹⁹ Widely applied for the market return are stock market indexes like the Dow Jones' Standard & Poor's 500 for the U.S.¹²⁰

The last missing and in literature most frequently discussed parameter for calculating the cost of equity is the beta. It provides information about the project's risk compared to the market portfolio's risk, also called systematic risk. A straightforward way of calculating beta starts with computing the covariance of the returns of a project i with the market portfolio m , Cov_{im} . In the next step this covariance is divided by the variance of the market portfolio σ_m^2 as displayed in (4).¹²¹

$$\beta_i = \frac{Cov_{im}}{\sigma_m^2} \quad (4)$$

Another way of calculating beta is by a linear regression. When regressing the asset's returns against the market portfolio's returns, beta represents the slope of linear regression.¹²² The resulting beta will be higher than one for assets that are riskier, and below one for assets that are less risky than the market portfolio average. This implies that a project with a beta of one bears the same risk as the market portfolio and that a risk-less asset has a beta of zero.¹²³ However, this way of calculating beta uses historical returns which are not publicly available for private companies.

¹⁰⁸Cf. Damodaran (2006), p. 52.

¹⁰⁹Cf. Rosenbaum and Pearl (2013), p. 152.

¹¹⁰Cf. Noothout et al. (2016), pp. 41-42.

¹¹¹Cf. Modigliani and Miller (1958), pp. 295-296.

¹¹²Cf. Damodaran (2006), p. 612.

¹¹³Cf. Rosenbaum and Pearl (2013), p. 152.

¹¹⁴ Cf. Noothout et al. (2016), p. 41.

¹¹⁵Cf. Noothout et al. (2016), p. 23.

¹¹⁶Cf. Damodaran (2006), p. 28.

¹¹⁷Based on Fama and French (2004) and Rosenbaum and Pearl (2013), p. 154.

¹¹⁸Cf. Damodaran (2006), p. 35.

¹¹⁹Cf. Damodaran (2006), p. 38.

¹²⁰Cf. Rosenbaum and Pearl (2013), p. 155.

¹²¹Cf. Damodaran (2006), p. 32.

¹²²Cf. Damodaran (2006), p. 48.

¹²³Cf. Rosenbaum and Pearl (2013), p. 156.

The fact that this calculation of beta only takes the volatility of the returns into account has led to further research on methods for determining the cost of equity. As an example, the Fama-French three-factor-model incorporates the market capitalization and the book-to-market ratio to the stock's or in this case project's returns.¹²⁴ Nevertheless, for this thesis it is not necessary to dig deeper into calculation methods for cost of equity as the introduced theory already provides clear evidence that they are significantly influenced by volatile revenues. The level of cost of equity for onshore wind projects differs substantially across Europe from very low rates between 6 % and 9 % in Germany up to 16 % to 18 % in Romania or 15 % to 20 % in Estonia.¹²⁵

The other share of the WACC apart from the cost of equity is the cost of debt. The latter quantifies the interest rate banks or private lenders demand for borrowing money. As abovementioned, debt is the cheaper type of financing compared to equity since debt lenders are served first in the case of financial failure.¹²⁶ For calculating the cost of debt, there is no commonly used method like the CAPM. In general, the cost of debt depends on a company's default risk. The latter is determined by the height of the project's generated cash flows in relation to its obligations and the stability of these cash flows. This implies that companies with high financial obligations and low but volatile cash flows end up having a high default risk. Due to this fact, high stability and predictability of the business a corporation operates in influence the default risk in a positive way. The default risk can be measured most easily if the company has outstanding bonds that are rated by independent rating agencies.¹²⁷ In general, the cost of debt is the sum of the risk-free rate and a credit risk rate representing a default premium.¹²⁸ Often the cost of debt and therefore the height of the interest the lender asks for is assessed qualitatively by consulting debt capital market specialists.¹²⁹ Investors name three main influencing factors on cost of debt, namely the investment risk for the specific industry, general country risk and the degree of competition between debtors.¹³⁰ Another factor significantly influencing the cost of debt is the competition between potential creditors for particularly attractive projects. This can e.g. be a RES project in a country like Germany. The cost of debt for the example of onshore wind investments varies substantially within Europe from countries like Greece where lenders demand between 8.5 % and 12.5 % to Germany where very low interest rates from 1.8 % to 3.2 % prevail.¹³¹

However, the introduced RES support schemes like FITs tend to lever out market risk almost completely leading to the perception of RES projects as secure long-term invest-

ments.¹³² Nevertheless, with elimination or reduction of governmental support the exposure to volatile electricity prices increases. In combination with the intermittency of the output, proper predictions on future returns probably become almost unmanageable. The resulting volatility of the expected cash flows represents an investment risk. According to the above introduced factors influencing the cost of capital, this volatility and uncertainty of revenues is reflected by a higher cost of capital.

7. Economic feasibility case study

In this section, a case study is conducted to analyze the influence an energy storage has on the investment risk for wind energy projects. The following model is developed to investigate the application of an energy storage for a case where no subsidies for RES are provided. This would correspond to a post support scenario. The aim is to test if the hypothesis that a battery can reduce the investment risk of a wind park holds true for the chosen case. Furthermore, threshold values that allow an economically feasible operation of such an VPP are identified.

7.1. Methodology

For the task of this analysis, a stochastic or a deterministic model can be used. For deterministic models, the future is predictable and the result is reproducible due to predetermined input values. Stochastic models however are based on probability theory. Statistics build the input values and the results are distributions, consequently not reproducible.¹³³ The following base model is implemented in a deterministic way since this makes it easier to understand compared to a stochastic model. Hence, this model is more likely to gain wider acceptance, especially for practical application.¹³⁴ As a consequence of its complexity, stochastic modeling requires increased computational effort, therefore advanced software. Nevertheless, Scenario testing in combination with the deterministic model includes uncertainties, provides a distribution instead of an absolute value as a result and thus will allow risk assessment.

The present model consists of the four steps displayed in figure 5. At first, the operation of a wind-battery VPP is optimized regarding cash flows. Secondly, Monte Carlo scenarios are generated and evaluated. Subsequently, risk figures of the operation with and without a storage are compared. In the fourth and final step, the sensitivities are checked and critical values are identified at which the operation of such an VPP would become profitable.

To go into detail, in the first step, the operation of the VPP is optimized in a risk-neutral way. The optimization aims at maximizing the cash flows. For this purpose, constraints like

¹²⁴Cf. Fama and French (1992), p. 451.

¹²⁵Cf. Noothout et al. (2016), p. 44.

¹²⁶Cf. Damodaran (2006), p. 612.

¹²⁷Cf. Damodaran (2006), p. 64.

¹²⁸Cf. Bishop and Officer (2013), p. 5.

¹²⁹Cf. Rosenbaum and Pearl (2013), pp. 153-154.

¹³⁰Cf. Noothout et al. (2016), p. 3.

¹³¹Cf. Noothout et al. (2016), pp. 40, 43.

¹³²Cf. Noothout et al. (2016), p. 43.

¹³³Cf. Rottmann et al. (n.d.).

¹³⁴Cf. Cummins et al. (1999), p. 424.

¹³⁵Own figure.

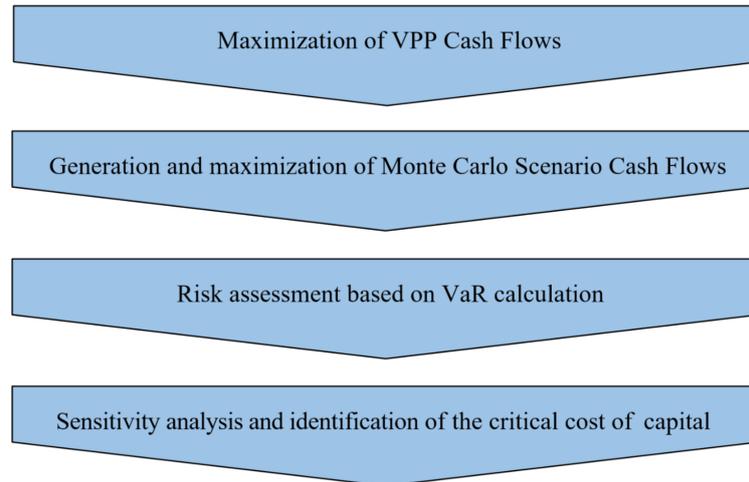


Figure 5: Steps of the developed model¹³⁵

capacity, power and efficiency degree of the battery must be defined. The model uses wind park output and electricity price data to develop a strategy, when to store and when to sell the produced energy directly. In order to optimize the operation of the VPP, a time horizon must be set. Fluctuations in prices and wind yield between different times of the day, workdays and weekends, different seasons and even between years exist. However, strategies for long time horizons like smoothening seasonal price and wind yield fluctuations would require huge capacities. This tends to be a case for large long-term storage capabilities like large hydro storages. As this thesis analyzes the impact of a battery storage in combination with wind, one day in terms of 24 hours was chosen as time horizon. This is due to the characteristics of battery technologies that include full power durations in this application below 24 hours.¹³⁶

The objective function of this optimization is stated in (5). P_d is the power which is directly sold, when produced, whereas P_{st} is the power sold from the storage. p_{spot} represents the spot price at the current time t . The objective is to maximize the sum of this function over 24 hours.

$$\sum_{t=0} (P_d(t) + P_{st}(t)) * p_{spot}(t) \xrightarrow{!} \max \quad (5)$$

For the objective function certain restrictions apply. Obviously, the amount of energy taken from the battery cannot be bigger than the amount already stored up to this time. Furthermore, P_{st} is restricted by the battery characteristics like capacity and power. Applying these restrictions, the model still remains idealized. However, the lifespan of a battery is subject to the number of charging cycles. Furthermore, losses occur when charging and discharging the battery. Consequently, imputed costs arise when using the storage. To de-idealize the operation of the storage and to consider the

aforementioned losses, the model includes a degree of efficiency for the battery. The side condition (6) holds true for every point in time. The maximum amount of energy that can be sold from the storage P_{st} equals the amount of energy stored at that time. This obviously cannot be higher than the difference of the wind turbine's output P and the amount of directly sold energy. This difference is multiplied with the degree of efficiency η . The level of the latter depends on the applied battery technology.

$$P_{st} \leq (P - P_d) * \eta \quad (6)$$

If the price difference is not big enough to compensate the $1 - \eta$ loss due to storing energy, the decision not to store energy at low prices and sell it at higher prices should be taken.

The wind turbine loads the battery when the current spot price is low. Consequently, the storage is discharged at a time with high electricity prices. This leads to a shift of the feed-in profile of the wind farm. Therefore, higher revenues result without increasing the total amount of energy produced. For the practical implementation of this optimization, Microsoft Excel Solver is used.¹³⁷

The cash flows of this VPP will then be compared to those of the standalone wind farm that must sell the energy at time and current price when produced. Another constraint is the specification that the load of the battery must be zero at the end of the day. This makes the implementation of the model easier since the daily operation schedule of the VPP is optimized. In this way, different days in terms of scenarios are more comparable.

In a second step uncertainty in terms of volatile output, fluctuating market prices and forecast errors is included into

¹³⁶Cf. Barton et al. (2004), p. 442.

¹³⁷Microsoft Excel Solver is an Excel add-in which can use different algorithms to find optimal solutions. In this application, the Generalized Reduced Gradient (GRG) non-linear algorithm is used. This is since some of the restrictions in the model do not meet the conditions for linearity, hence the Simplex algorithm could not be used (Cf. Frontline Systems, Inc. (2017)).

the model via scenario testing. Accordingly, spot price and wind scenarios are simulated using Monte Carlo method. Historical data used only projects and therefore estimates an expected value. The Monte Carlo simulation generates a range of possible outcomes.¹³⁸ Appropriate random numbers which are generated by using Monte Carlo simulation, serve as input values. For generating these random numbers, the forecast error for wind and spot prices is considered. The distributions of the latter are assumed to be normally distributed with zero mean.¹³⁹ To implement the Monte Carlo simulation RiskAMP was used.¹⁴⁰ Apart from the type of distribution, the mean and the standard deviation are required to generate a Monte Carlo scenario. For generating the scenarios for the energy output of the wind farm, the wind forecast error was used as the distribution's standard deviation. The forecast error δ_t in (7) is calculated as a function of the forecasted power P_f and the maximum output of the wind farm P_{max} .¹⁴¹

$$\delta_t = \frac{P_f}{5} + \frac{P_{max}}{50} \quad (7)$$

Similar to the wind forecast, the error for the price forecast depends on the forecast technique used. For the according scenarios, a deviation of 5 % of the estimated value is assumed.¹⁴² The simulated scenarios are also used as input data for the VPP optimization model like it is done with the real data in the first step of the model. The resulting yearly cash flow scenarios show a distribution for each, the VPP and the sole wind farm. They are used to assess the risk and a comparison of the operation with and without a battery.

The risk assessment represents step three of the model. To determine if there is an influence of using a battery storage in combination with a wind farm on the investment risk, different risk measures can be used. A nowadays commonly applied measure is the VaR. It provides information about the maximum loss that will not be exceeded in a certain time horizon at a defined confidence level. This equals the relevant distribution's quantile in terms of statistics.¹⁴³

In this application of the VaR, I do not investigate the maximum loss that could occur. Instead, it is applied to determine the minimum cash flow that will be realized with the specific probability in each period. The VaR is calculated at the 99 %, the 95 % and the 90 % confidence level. Consistency in the results for all three confidence intervals will increase the results' validity. If the VPP achieves a higher minimum cash flow relatively to the expected value compared to the standalone wind farm, the use of a battery is likely to have a positive effect on the investment risk. An improved VaR corresponds with a reduction of uncertainty coming from

different price and output scenarios. This implies a narrower distribution of the cash flows for the VPP compared to the cash flows of the standalone wind farm for the same scenarios.

Given that the VaR analysis states a lower risk for the VPP compared to the standalone wind farm, the fourth step follows. It is assumed that with a reduced risk, the WACC of the investment decreases. This complies with the assumption that uncertainty and volatility in revenues influences both, the cost of debt and the cost of equity. Section 5 elaborates on this topic. Consequently, the WACC is used as total cost of capital. A further assumption is that the innovation character of combining a wind power plant with a battery storage in a VPP is not influencing the investment risk and the cost of capital. The implementation of the fourth step starts with the calculation of the wind farm project's NPV by using the DCF method. Therefore, a reasonable duration of operation and WACC have to be determined. The resulting NPV is compared to the NPV of the VPP for which the initial investment of the battery must be considered. The price of the battery storage depends on its capacity on the one hand and the price per kWh for the chosen technology on the other hand. The initial investment of the wind farm itself is not taken into consideration as the VPP is compared to the standalone wind farm. Consequently, when taking this investment into account, results remain unchanged. For the wind power plant without a battery, no initial investment at all is included into the NPV. Subsequently, the critical WACC for the VPP is determined. The latter corresponds to the discount rate, at which the NPV of the VPP matches the NPV of the standalone wind farm. The calculation is also conducted by using the Excel Solver which is capable of performing this computation by applying what-if analyses. Since the cash flows are optimized and consequently set, the height of the critical WACC depends on the investment cost of the battery and the defined timeframe. For investigation of these dependencies, sensitivity analyses with battery price and number of periods as varying inputs are conducted. In this context, the threshold battery price, for which the NPVs would be equal for the VPP and the standalone wind farm is identified, assuming the same WACC.

7.2. Data set

For the study, the output pattern of an existing representative 50 MW wind farm in northern Germany is considered. As data basis, the output data for the years 2013 – 2015 is available. This selected data sample is favorable for the study as 2013 was a poor, 2014 a medium and 2015 a good year in Germany when it comes to overall wind yields.¹⁴⁴ Hence, bias coming from the choice of a specific year is mitigated. The plant's wind yield data is provided as quarter-hourly values. These outputs are consolidated to hourly values to match the time interval of the available price data. As a next step, the data is pre-processed to eliminate irregularities and make it easier to handle via generating typical profile days.

¹³⁸Cf. Structured Data LLC. (n.d.).

¹³⁹Cf. Liu et al. (2015), p. 785.

¹⁴⁰RiskAMP is an add-in for Microsoft Excel which provides a Monte Carlo simulation engine (Cf. Structured Data, LLC. (2017)).

¹⁴¹Cf. Liu et al. (2015), p. 785.

¹⁴²Cf. Conejo et al. (2005), p. 1039 f.

¹⁴³Cf. Jorion (2006), p. 17.

¹⁴⁴Cf. IWR (2014); IWR (2015); IWR (2016).

This consolidation of data is common practice in research and simplifies the handling of bulk data.¹⁴⁵ Therefore, an average day is generated for each month of a year. This is done as wind levels are not constant over the year. The result as can be seen in figure 6, shows that the output in winter months is much higher than in summer. For example, the average output on a December day is more than 390 MWh whereas on an average august day only around 103 MWh are produced.

For electricity price data, the European Power Exchange (EPEX) spot prices for Germany are used. The prices which were available on an hourly basis over the same three years are also broken down to typical daily patterns. This is necessary as the prices vary between months. However, for electricity prices it is insufficient to only distinguish between months. The key influence on the height of the prices over the day, and thus on how the load profile of a day looks like, is the fact if it is a working day or weekend.¹⁴⁷ Daily load profiles of weekends are very different to those of working days as can be seen in figure 7. In detail, typical price trends of January working days and weekends are compared. Figure 8 displays the average prices per MWh for the two categories of weekdays over the year.

A major role in creating the introduced VPP is the battery capacity as well as the maximum power of the storage. For the first step of the analysis, the cash flow maximization, an endlessly big storage would be the best, as all the generated electricity could be sold at the time with the highest price and investment costs would be not considered in this part of the model. Nevertheless, investment cost is a substantial issue and choosing the right capacity and power is crucial for maximizing the benefit. The topic of optimal sizing of batteries for VPP is an issue that makes up an own field of research. For this thesis, the size of the storage is determined in a pragmatic way. As the storage is used to maximize the daily profit, the average daily output of the wind farm is the first step of figuring out an appropriate capacity. The daily average output over the years 2013 until 2015 is about 198 MWh, whereas the hourly average output is 8.3 MWh.

For this application, the battery has to shift the output within one day. Therefore, the load change has to be quite fast and high power has to be provided. Furthermore, the used battery should have a long lifespan. For this case a lithium ion battery is chosen. To use a storage with half of the days average output would result in a battery system with almost 100 MWh. This nearly equals the size of the currently installed world's largest lithium ion battery storage in California that has a capacity of 120 MWh.¹⁵⁰ The ratio between capacity and wind farm output in practical applications differs significantly. The Notrees¹⁵¹ MW wind farm in Texas for

instance was equipped with a 24 MWh and 36 MW battery which results in a capacity to wind power ratio of around 0.16.¹⁵³ Another VPP project in Braderup in northern Germany consists of an 18 MW wind farm and a 3.3 MWh Battery.¹⁵² In a study on a VPP consisting of wind power and Vehicle-to-grid energy storing, storage capacities from 19 to 50 MWh are analyzed for a 13 MW wind farm.¹⁵³ The relation of storage size to wind farm output varies for different applications as well. It makes a difference, if the objective of the storage is to maximize revenues, to stabilize outputs or to participate in the balancing energy market.¹⁵⁴

For this study, a 25 MWh lithium ion battery system was chosen. On the one hand, this is a storage for today's conditions. On the other hand, the fact that this model is developed for the operation without financial support which is rather a case for after 2020, battery prices are likely to be lower. Therefore, this is assumed to be a realistic dimension. For the power of the battery 25 MW is chosen. As battery prices are usually provided in € or \$ per MWh, the maximum power as price factor is neglected. For this model, it is assumed, that the lithium ion cells are available as units with a capacity that allows one hour of discharging with the maximum power and hence a ratio of power to capacity of 1. This is assumed to be a good starting point, when considering the application of second-use electric vehicles batteries since these usually have rather high power. To make a statement about the influence of the battery size, the study is additionally conducted with a 15 MWh and a 35 MWh battery. This can provide information about additional potential of risk reduction and about an increase in profitability through variations in battery capacity.

As degree of efficiency for the battery 90 % are used.¹⁵⁵ Degrees of efficiency can be better nowadays and might, because of further research and development in this field, be even better for the time, when the assumed circumstances of the model exist.¹⁵⁶ This may be from the year 2020 onwards. Nevertheless, 90 % are used as this also considers the cost of wear and tear by charging and discharging the battery in this study. Furthermore, the degree of efficiency probably will be lower for the application of second-life electric vehicle batteries.

To create the Monte Carlo Scenarios random values are generated using the RiskAMP Excel Add-In. As abovementioned, the prediction errors for both, the wind and the spot prices are assumed to be normally distributed with zero mean. In total, 10 scenarios for each typical day are generated and the daily revenues are maximized for each of them. This number of scenarios was chosen to keep computational and manual effort manageable. Figure 9 shows the Monte Carlos price scenarios for working days in January. The red line which represents the typical day derived from the real

¹⁴⁵Cf. Hippert et al. (2001), p. 49.

¹⁴⁶Own figure based on wind park data 2013-2015.

¹⁴⁷Cf. Hippert et al. (2001), p. 49.

¹⁴⁸Own figure; based on the EPEX Spot Day-Ahead prices for 2013 – 2015.

¹⁴⁹Own figure; based on the EPEX Spot Day-Ahead prices for 2013 – 2015.

¹⁵⁰Cf. Overton (2017).

¹⁵¹Cf. Better World Solutions (2015); Younicos (2016).

¹⁵²Cf. Gillhuber (2014).

¹⁵³Cf. Vasirani et al. (2013), p. 1320.

¹⁵⁴Cf. Doetsch et al. (2011), p. 100.

¹⁵⁵Cf. Statista (n.d.).

¹⁵⁶Cf. Leuthner (2013), p. 16.

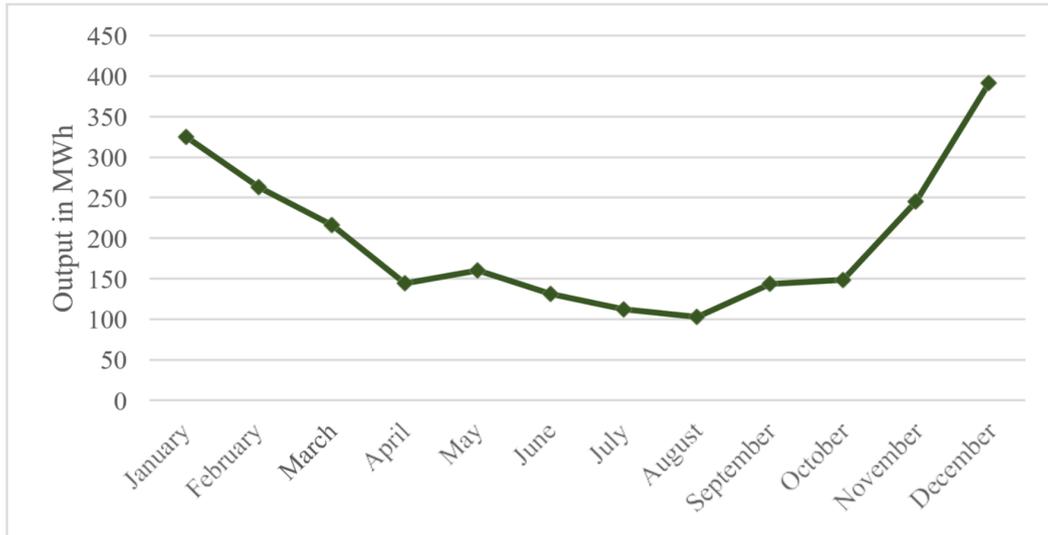


Figure 6: Average daily wind farm output¹⁴⁶

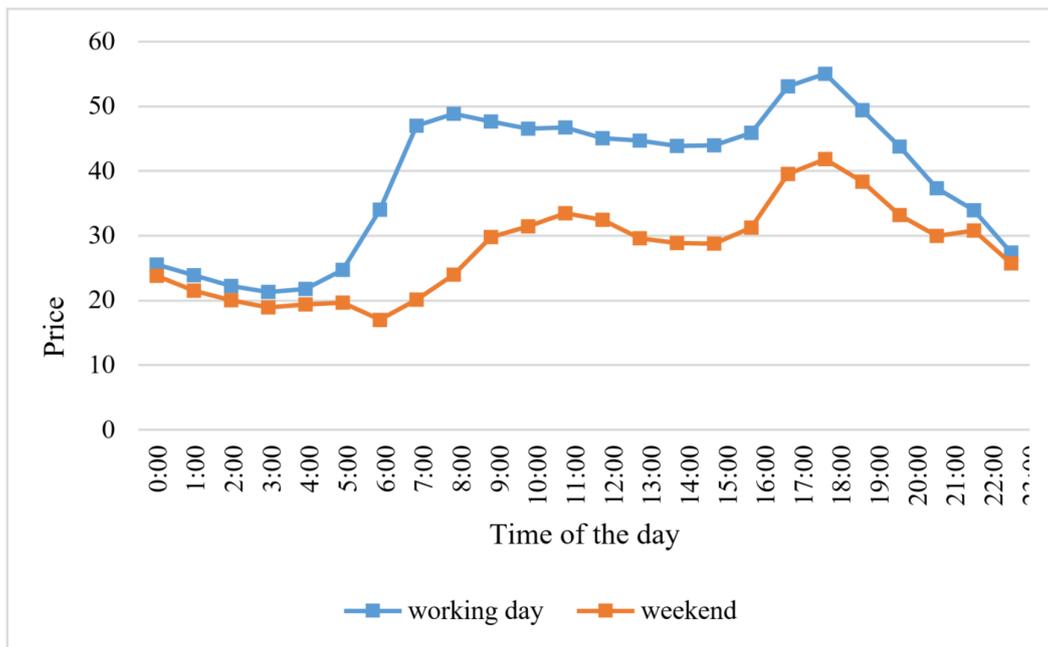


Figure 7: Typical daily price profile for January¹⁴⁸

price data therefore equals the blue line in figure 7. Figure 10 includes the output scenarios of the wind farm in a structure analogous to figure 9. Worth mentioning is the fact that the output pattern over a day usually will not be as smooth as the red line as this line represents a typical day calculated as average.

For the calculation of the VaR a bigger number of scenarios is needed, since e.g. the 99 % VaR equals the 1 % quantile. For only 10 values this requires an interpolation and might

end in an unprecise estimation. The operation for the generated Monte Carlo scenarios is optimized for each of the 10 cases. The accompanying monthly revenues are independent of each other. This independency is used to create a bigger sample by combining the monthly outcomes of the optimization of the scenarios to new scenario years. This is done since the VaR will be calculated for the yearly cash flows. If every month would be combined with each other, this would result in 1012 combinations. This number of scenarios however would exceed the capabilities of Excel. To enlarge the sample by choosing some scenario combinations without getting a biased selection of combinations, 500 random combi-

¹⁵⁷Own diagram.

¹⁵⁸Own diagram.

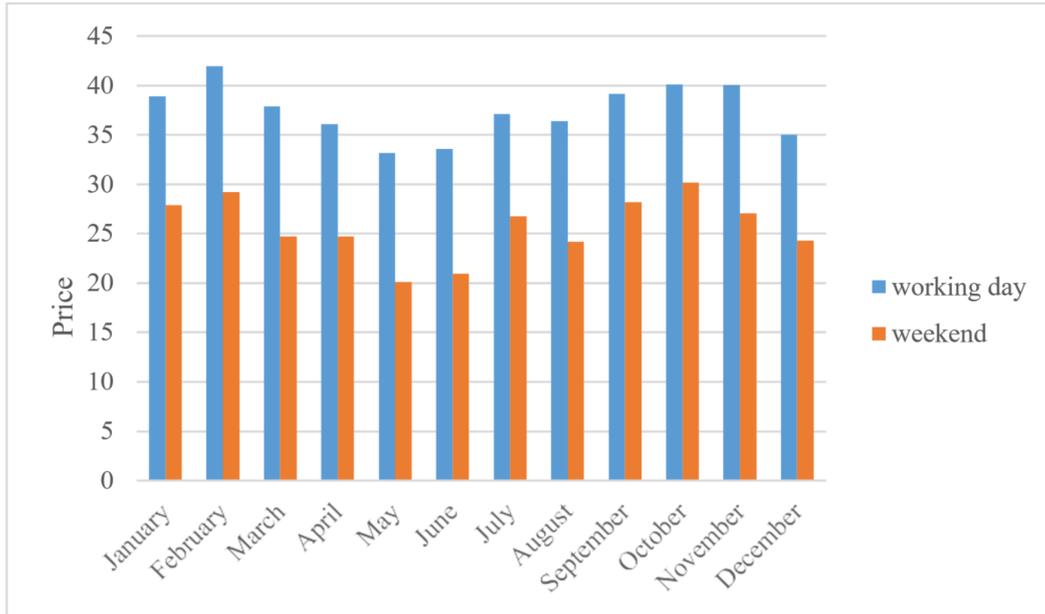


Figure 8: Average electricity price on typical days¹⁴⁹

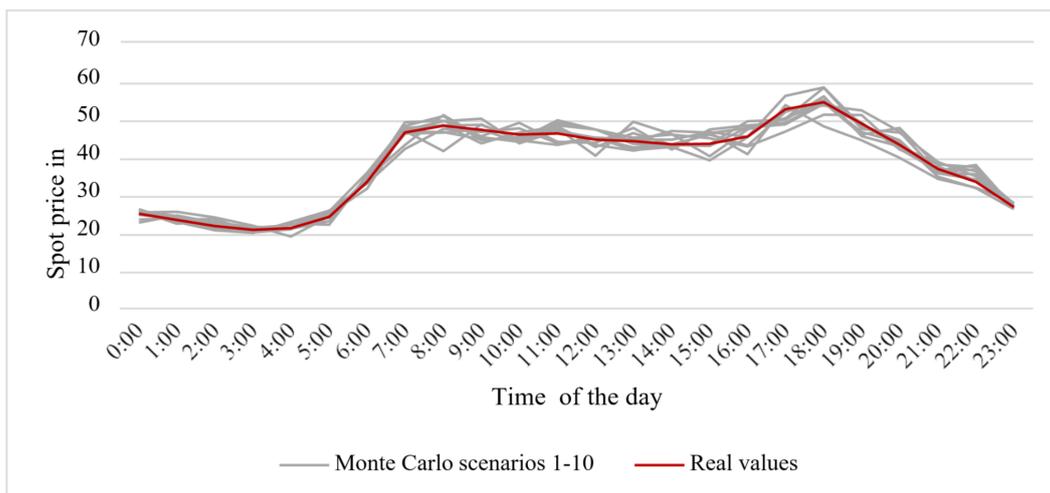


Figure 9: Price scenarios for a working day in January¹⁵⁷

nations are generated. For conducting the VaR calculation, confidence levels of 90 %, 95 % and 99 % on the annual cash flows are chosen. Due to the fact that the VaR is calculated on the revenue distribution over the different scenarios, the implication from this study is not the influence a battery has on daily volatility, but on how it can reduce deviations from the expected value for different scenarios. As in this case the VPP consisting of this plant and a battery is compared with the standalone wind farm, the investment cost of the wind turbines is of secondary importance, whereas the investment cost of the battery storage unit is of greater importance.

Since the case considers a post support scenario, it will take place e.g. from 2020 onwards for the case of Germany. This implies that prices for lithium ion batteries will considerably differ to today's prices because prices tend to be falling

for this technology. Studies estimate that prices in 2020 will decrease to 200 € per kWh and evolve to prices of around 160 € per kWh in 2025.¹⁵⁹ Primarily, the mass production as a consequence of a higher demand for EVs is the reason for this forecasted price decline. Cases with both, 200 € and 160 € are assessed.

The increasing share of EVs also leads to a growing availability of second use batteries as mentioned in section 2.4. The prices of the latter might be even lower and would possibly increase the project's profitability.

To calculate the NPV of the project, a WACC is needed as discount rate. In order to figure out an appropriate WACC,

¹⁵⁹200 \$/kWh was cleared with the up-to-date exchange rate (1 USD = 0.9421 €) (Cf. D'Aprile et al. (2016)).

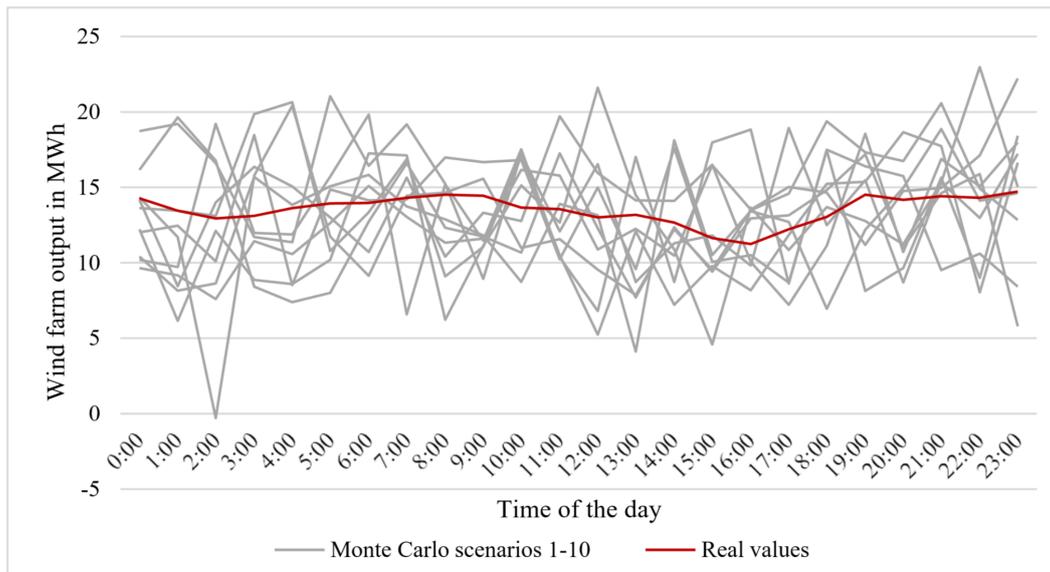


Figure 10: Output scenarios for a day in January¹⁵⁸

the DiaCore project is used. For Germany, which is the considered country as the wind farm is located there, the model value of the study is 5.6 %, whereas the estimations from the interviewees indicate a WACC in a range from 3.4 % to 4.5 %.¹⁶⁰ Within the scope of the initial investment, 5 % are used and for sensitivity testing, the NPV is recalculated with 3.5 % and 6.5 %. Another required input for the NPV calculation is the number of periods. For the project's time horizon, a 20-year timeframe is chosen. For sensitivity-testing, the calculation is conducted with a 15- and a 10-year horizon as well. The critical storage cost where the NPVs with the same timeframe and WACC match for the standalone wind farm and the VPP was calculated for the mentioned horizons and discount rates reaching from 3 % to 7 %.

7.3. Results

The results of the first step of the analysis match the expectations as a storage is for the VPP used which generates additional revenues by shifting outputs to high price times. Still, the investment cost of this storage unit is not yet included in this part. Figure 11 shows the average over the monthly revenues of the real data and the generated Monte Carlos simulations for the standalone wind farm and for the optimized operation of the VPP with a 15 MWh battery. The fact that the revenues of the VPP are higher is as expected. The profitability of the battery depends on the financing parameters like initial investment cost and the cost of capital which partly depends on the risk of the investment.

The VaR of the standalone wind farm and the VPP that is calculated by using the 500 generated random combinations of the scenarios are shown for the different battery capacities in table 3.

The absolute number can be interpreted as a minimum yearly revenue with a certain probability. Thus, with a 90 % probability the yearly revenue of the VPP with the 25 MWh will not be below 2,672,830 € whereas it will not fall below 2,394,245 € for the standalone wind farm for the same wind and price scenarios. The relative shortfall from the expected value that equals -2.29 % for the VPP is smaller than for operation without a battery which equals -2.50 %. To conclude, the VaR calculation states for the 90 %, the 95 % and the 99 % confidence level that the operation of the VPP is less risky than operating the equivalent wind farm without the possibility of storing energy. The calculation with all three confidence levels increases the validity of the analysis.

Especially interesting in the VaR analysis is the influence of the battery size on the risk. Obviously, the absolute numbers for the VaR of the 35 MWh battery in terms of the quantiles are higher because the whole distribution of the revenues is on a higher level compared to smaller batteries. Nevertheless, the relative shortfall from the mean is bigger than for the 25 MWh battery for all three confidence levels. A possible reason for this phenomenon is the fact that in seasons with a low wind yield, the capacity remains partly unused due to the large-scale of the storage. The conclusion from this assessment and the highlighted cells in table 3 is that the 25 MWh battery could reduce risk better than a 15 MWh and 35 MWh battery.

The reduction of the risk of wind farm investments in terms of lower volatility of future cash flows is assumed to cause a reduction of the WACC. The following table 4 illustrates the critical values for the WACC. The latter must lower for the VPP to this value to achieve the same NPV as the standalone wind farm does. For example, when there is a WACC of 5 % for wind power investments, a battery price of 200 € /kWh and a project duration of 20 years is assumed, a WACC of 4.46 % and therefore a reduction by 0.54 per-

¹⁶⁰Cf. Noothout et al. (2016), p. 112.

¹⁶¹Own figure.

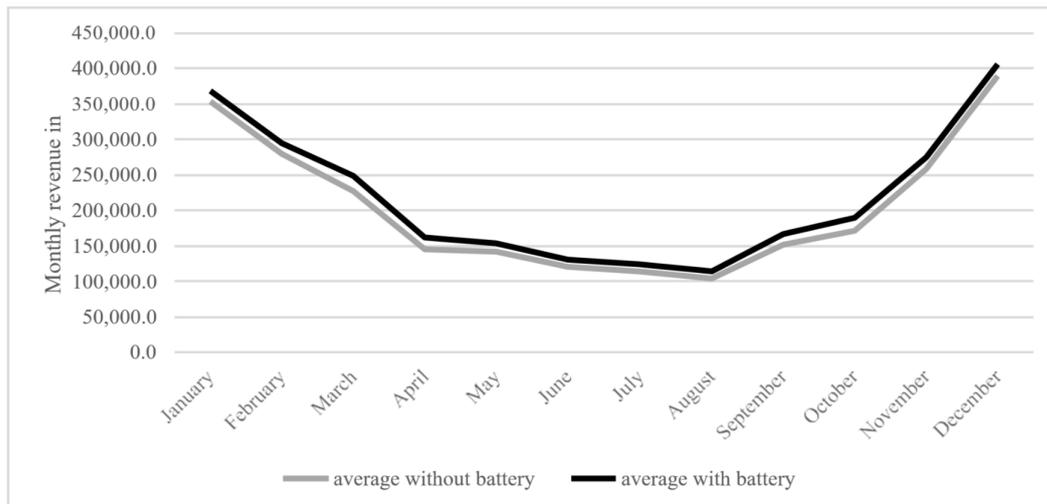


Figure 11: Average monthly revenues with and without battery storage¹⁶¹

Table 3: Value-at-Risk results based on yearly revenues; Source: Based on own calculations.

VaR	Standalone wind farm		VPP 15 MWh		VPP 25 MWh		VPP 35 MWh	
	Relative	Absolute number	Relative	Absolute number	Relative	Absolute number	Relative	Absolute number
90 %	-2.50 %	2,394,245 €	-2.38 %	2,570,291 €	-2.29 %	2,672,830 €	-2.30 %	2,749,306 €
95 %	-3.38 %	2,372,719 €	-3.18 %	2,549,210 €	-3.03 %	2,652,676 €	-3.06 %	2,727,944 €
99 %	-4.75 %	2,339,055 €	-4.37 %	2,517,781 €	-4.12 %	2,622,987 €	-4.15 %	2,697,407 €

centage points is necessary for the 25 MWh battery VPP as a threshold value to match the NPV of the windfarm without battery. In case of a cost of capital rate below this critical value, the VPP has a higher NPV than the plant without energy storage. For a WACC of 3.5 % with a project’s timeframe of 20 years and a battery price of 160 € /kWh, which is reasonable for 2025, the VPP would already be almost competitive since the cost of capital would only have to fall to 3.46 %.¹⁶² Striking is that shortening the project’s lifetime from 20 to 15 years for a 160 € /kWh battery nearly has the same influence for the analyzed rates of cost of capital as raising battery prices from 160 € /kWh to 200 € /kWh. Also noticeable is that the shorter the time horizon is the bigger is the influence of higher battery prices. This seems reasonable as there are less years with higher cash flows

for the VPP compared to the sole wind farm which contribute to cover the initial battery investment. The 40 € /kWh difference in battery price sum up to a significant difference of 1 million € for the initial investment for the 25 MWh energy storage. As concluded from the VaR calculations, the 25 MWh battery can reduce risk by more than the 15 MWh and the 35 MWh battery. And since the 35 MWh battery would have to reduce the cost of capital more than the other configurations, this tends to be not an interesting

option. However, since the initial investment is lower, the 15 MWh would only require a smaller reduction of the WACC to match the NPV of the standalone wind farm. For the 15 MWh battery, there is even the case where the additional revenues would completely cover the initial investment of the storage. This underlies the assumption that the most favorable of the assessed combinations occurs, with the battery price being 160 € /kWh, the project’s period 20 years and the WACC for wind projects 3.5 %. Consequently, the NPVs of the standalone wind farm and the VPP would match without the necessity to lower the WACC for the VPP. However, there is evidence, that this is mainly caused by the favorable conditions since the WACC would only have to be reduced from 3.50 % to 3.46 % for the 25 MWh battery. As for this study there was merely the assumption that a better VaR reduces the WACC, it would be an interesting extension for future research to quantify the impact a certain change in the VaR has on the cost of capital.

The results of the calculations about the critical battery prices which directly can be projected to the critical initial investment, can be seen in figure 12. The previous calculation of the critical WACC obviously indicates that a longer project’s life allows higher prices. As an example, the critical energy storage price for a 25 MWh battery at which the NPVs for the VPP and the standalone plant are equal, at a WACC of 5 %, is 136 € /kWh for a 20-year, 113 € /kWh for

¹⁶²Cf. D’Aprile et al. (2016).

Table 4: Critical WACC for the VPP; Source: Own table; An overview of the calculations and results for one battery configuration (25 MWh) is provided in appendix 1.

Battery capacity	Battery price					
	160 € / kWh			200 € / kWh		
	Timeframe					
	10 years	15 years	20 years	10 years	15 years	20 years
	Reference WACC = 3.5 %					
15 MWh	2.63 %	3.30 %	3.50 %	2.13 %	3.03 %	3.33 %
25 MWh	2.07 %	3.12 %	3.46 %	1.31 %	2.70 %	3.19 %
35 MWh	1.37 %	2.82 %	3.31 %	0.39 %	2.27 %	2.94 %
	Reference WACC = 5.0 %					
15 MWh	3.94 %	4.66 %	4.90 %	3.39 %	4.35 %	4.69 %
25 MWh	3.26 %	4.41 %	4.79 %	2.43 %	3.93 %	4.46 %
35 MWh	2.47 %	4.03 %	4.58 %	1.41 %	3.40 %	4.14 %
	Reference WACC = 6.5 %					
15 MWh	5.23 %	6.01 %	6.27 %	4.62 %	5.65 %	6.02 %
25 MWh	4.44 %	5.66 %	6.09 %	3.53 %	5.12 %	5.70 %
35 MWh	3.54 %	5.20 %	5.80 %	2.40 %	4.50 %	5.30 %

a 15-year and 84 € /kWh for a 10-year time horizon. This is due to the fact that the additionally generated revenue by deploying the battery needs to cover its initial investment. With a higher number of years in which these additional revenues occur, more overall cash flows cumulate. However, the figures that are further away in the future contribute less to covering of the cost than the early returns due to discounting.

In figure 12, it is noticeable that the lines seem to be flatter for shorter project periods. Consequently, the critical battery prices for a 10-year project are obviously lower than for longer projects but the influence of higher cost of capital also seems to be lower. Furthermore, the difference in threshold values for batteries is bigger between 25 MWh and 35 MWh than between 15 MWh and 25 MWh.

All in all, it can be stated that for projects with 15 to 20 years, the application of a battery storage in a VPP can be competitive to a standalone wind farm. This assumes that in the future battery prices will be significantly lower than today or where second-use storages from electric vehicles can be used. Notably, this holds for a scenario without any support scheme for renewables. Furthermore, it depends on the development of the WACC for RES investments. Even if there is a degree of efficiency for the battery in the model, this still would not be sufficient to represent a real, non-idealized battery. Including more parameters like a lower partial load limit, operating cost and cycle stability would result in changes of the optimal operation schedule and the economic performance.¹⁶⁴ Additionally, the assumption that the standalone wind farm is self-balancing and sells the energy

produced without a third-party wholesaler results in higher cash flows. When considering a direct marketer charging a predefined share from the standalone wind farm operator, the difference between the revenues compared to the VPP clearly will increase.

8. Alternative instruments for revenue stabilization

Even though this thesis emphasizes the combination of RES with energy storage to deal with the challenges of the electricity market in a post-2020 scenario, there are other instruments to successfully integrate RES into the market. Figure 13 shows the classification of alternative risk stabilizing instruments covered in this chapter. Firstly, there is the possibility of transferring the risk of volatile electricity market prices and intermittent output by directly selling electricity to consumers in terms of power purchase agreements (PPA), using energy derivatives or by using insurances. Besides transferring the risk, the possibility of adjustments on the market framework for RES dominated markets exist. Finally, geographic diversification as instrument to reduce output volatility is assessed.

8.1. Risk transfer

8.1.1. Corporate power purchase agreements

While average households buy power from utilities, some companies purchase directly from generators by long-term contracts called PPAs. These kind of contracts have been used almost exclusively for conventional power sources for

¹⁶³Own figure.

¹⁶⁴Cf. Doetsch et al. (2011), p. 89.

¹⁶⁵Own figure.

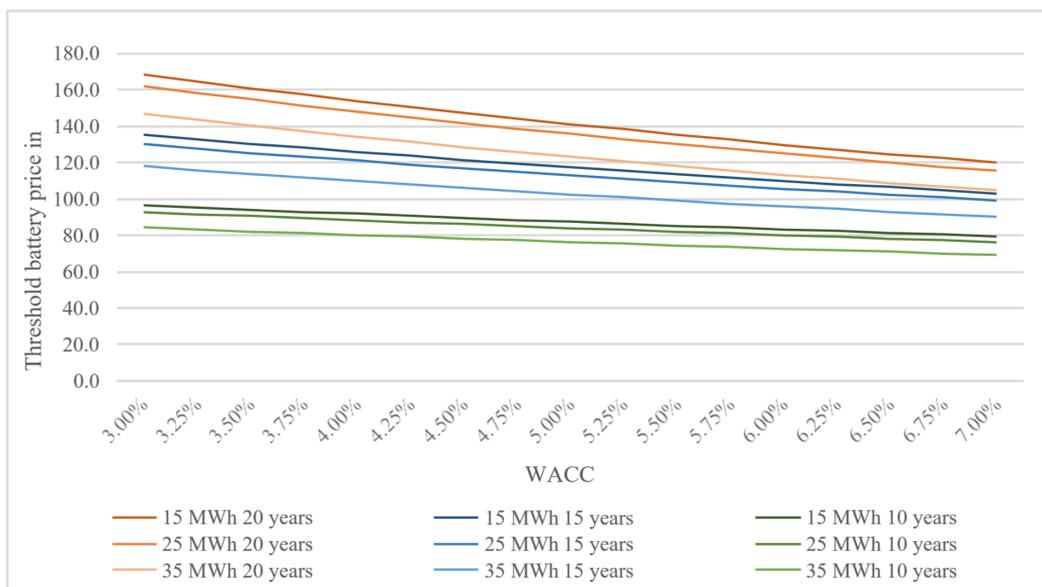


Figure 12: Critical battery prices dependent on WACC and timespan¹⁶³

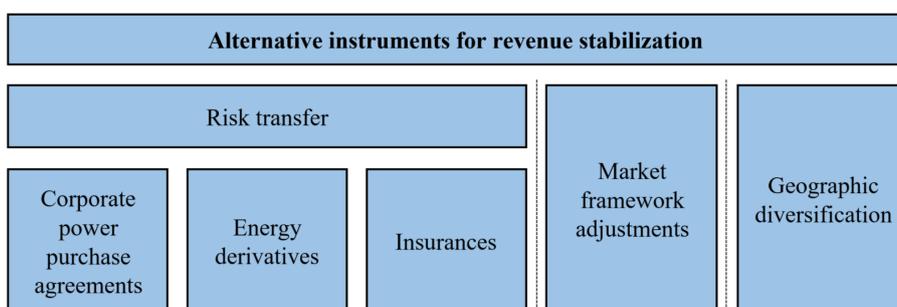


Figure 13: Structure of the section on alternative risk stabilizing instrument¹⁶⁵

a long time.¹⁶⁶ It is common practice that banks ask for a PPA as a prerequisite for providing debt.¹⁶⁷ PPAs are usually at least ten year term contracts between a company as electricity consumer and a renewable energy generator. The consumer usually commits to take all the produced energy of the plant for a fixed price per kWh. The delivery of the electricity can either be physical or virtual.¹⁶⁸

For physical PPAs the generated electricity purchased is delivered directly from the seller to the purchaser meaning the consumer is receiving the actually bought electricity.¹⁶⁹ Businesses with big data centers or other facilities that have a concentrated heavy load are common companies in physical PPAs. Usually in a physical PPA, the seller is responsible for providing the electricity at a particular delivery point, whereas the purchaser has the duty to transfer the electricity to its load. This service is often taken over by external service providers.¹⁷⁰

Virtual PPAs can be structured in different ways and are closer to a financial contract or an option than an actual contract for power. Virtual PPAs are hedging instruments to reduce the risk coming from volatile electricity prices. A common form of this kind of contract is the contract for difference (CFD). For a CFD the buyer and the seller agree on a fixed strike price for the electricity. Therefore, there is a guaranteed price for both parties regardless of the actual volatile market price. In a first step, the generator sells the energy to the market and receives the market price. The consuming business obtains the electricity from the utility as usual and pays the market price. For a market price exceeding the settled strike price of the CFD, the seller pays the difference to the buyer and vice versa. Using this kind of PPA no adaptations regarding the process of how consumers purchase the energy from the utility have to be made.¹⁷¹ Another advantage of virtual PPAs is the simplicity for logistics and the fact that the consuming business can virtually power more than one subsidiary by one plant.¹⁷²

¹⁶⁶Cf. Baker & McKenzie (2015), p. 2.

¹⁶⁷Cf. Green Rhino Energy (2013).

¹⁶⁸Cf. DLA Piper (2016), p. 6.

¹⁶⁹Cf. Baker & McKenzie (2015), p. 11.

¹⁷⁰Cf. Penndorf (2016).

¹⁷¹Cf. Baker & McKenzie (2015), p. 11; Penndorf (2016).

¹⁷²Cf. Baker & McKenzie (2015), p. 11.

For consumers this form of purchasing electricity is a hedge against price uncertainty, can help them to put emphasis on their sustainability agendas by reducing their carbon footprint and improves its public recognition of being an environmentally-friendly.¹⁷³ There is also an initiative of big businesses from all over the world called RE100 which committed themselves to 100 % renewable energy. At this moment, 89 big Corporations and electricity consumers like IKEA, BMW, Coca Cola, General Motors, Google, Procter & Gamble and many more are part of this initiative.¹⁷⁴ The advantage of PPAs for the plant operators is the fact that the fixed electricity price provides predictable and stable revenues which gives certainty and mitigates market risks like the volatile spot price. Besides the operator, the investors of the plant benefit from this market risk reduction. This possibility of removing market risk leads to new implemented projects in RES.¹⁷⁵

The amount of electricity purchased directly by companies via PPAs has significantly increased over the last years. Bloomberg New Energy Finance analyzes recent PPA deals frequently and lists the biggest corporate off-takers. For 2016 e.g. this was Amazon with 650 MW, followed by other big companies like Google, Microsoft, Facebook, Dow Chemical or Walmart. During this year, the amount of electricity purchased via PPAs tremendously rose compared to prior years.¹⁷⁶

Companies as consumers have capability to further push additional conduct of RES projects via PPAs as they are responsible for a large share of electricity. Nevertheless, corporate PPAs alone are not perceived to be able to initiate sufficient investments in order to reach the 2030 renewable energy targets.¹⁷⁷

8.1.2. Energy derivatives

Apart from corporate PPAs there are other derivatives to hedge market risk. A reason why derivatives as financial tools can be advantageous compared to PPAs is the fact that PPAs often come along with discounts resulting in lower prices compared to forward prices.¹⁷⁸ A special characteristic about trading electricity is the fact that there are restrictions on storage and transportation in contrast to other commodities.

A common derivative and primary instrument in electricity price risk hedging is the forward. Electricity forwards are bilateral contracts which preliminary are traded in the Over-the-Counter (OTC) market. A forward is an individual contract which consists of the obligation to sell or buy a specific underlying at a certain time in the future at a predetermined price. For electricity, the aforementioned underlying is a fixed amount of electricity. Consequently, electricity forwards are a type of supply contract which obliges the seller

to deliver and the buyer to take the electricity. Electricity forwards' maturities are in a range from a few hours to several years. Nevertheless, maturities longer than two years are unusual.

What differentiates electricity forwards from other commodity forwards is the distinction of different delivery times at a day. Consequently, electricity is a dissimilar commodity at different day-times. There is on-peak and off-peak electricity which is traded separately and delivered at different times of the day. Therefore, the average electricity price over the relevant time interval at maturity is typically used for calculating the settlement price. Usually, the electricity producer or plant operator is the seller in this type of contract, whereas a utility or a consuming company represents the buyer. Per definition the seller of the underlying holds the short position and the buyer the long position.

Aside from the deviating calculation of the settlement price, the general calculation of the payoff of an electricity forward is equal to other financial or commodity forwards. This is shown in (8)¹⁷⁹, where S_T represents the spot price at maturity T and F the pre-specified contract price.

$$\text{ForwardPayoff} = (S_T - F) \quad (8)$$

The forward payoff results in a profit for the buyer and a loss for the seller if the spot price at maturity exceeds the forward price and vice versa. One example for financially settled electricity forward contracts is the abovementioned CFD.¹⁸⁰

With the same structure as forward contracts, future contracts represent the highly standardized and exchange traded equivalent. While forwards are usually OTC traded and individualized, futures are standardized in several specifications like settlement procedures and trading locations. The traded delivery electricity quantity is usually substantially smaller for electricity futures than for forwards. Futures are only traded on organized exchanges unlike forwards. The result is higher transparency compared to forward prices and a reduced risk of default of contracts counterparty. The major drawback of electricity futures can be the limitation when it comes to transaction quantities.¹⁸¹

Another widely used derivative is the swap. A swap represents a contract for the exchange of financial instruments between two parties. The interest rate swap is the most applied form of a swap. It allows the exchange of a floating interest rate for a fixed income rate between two parties where one party holds a contract with a floating interest rate and the other party with a fixed interest rate. For specific reasons, like hedging against interest rate uncertainty or speculating in a specific change of floating interest rates, these parties can enter a swap contract.¹⁸²

The electricity swap represents an application in the electricity market. It allows the holder to exchange a floating

¹⁷³Cf. WindEurope (2017), p. 15; DLA Piper (2016), p. 5.

¹⁷⁴Cf. RE100 (n.d.).

¹⁷⁵Cf. Baker & McKenzie (2015), p. 6; WindEurope (2017), p. 15.

¹⁷⁶Cf. Bloomberg New Energy Finance (2016), p. 4.

¹⁷⁷Cf. WindEurope (2017), p. 16.

¹⁷⁸Cf. Aydin et al. (2017), p. 2.

¹⁷⁹Cf. Deng and Oren (2006), pp. 942-943.

¹⁸⁰Cf. Deng and Oren (2006), pp. 942-943.

¹⁸¹Cf. Deng and Oren (2006), pp. 943-944.

¹⁸²Cf. Investopedia (2017).

electricity price for a fixed price over the contracted period. Typically, the contract is defined for a fixed quantity of energy and the floating price is referenced to a spot price. Theoretically, swaps are a strip consisting of multiple forward contracts with a constant forward price for each of the multiple settlement dates.¹⁸³

Aydin et al. suggest a basic hedge structure based on forwards for wind plants exposed to market price risk.¹⁸⁴ The basic idea is exchanging volatile prices with a fixed price by selling forward contracts at expected wind output levels. The difficulty in this strategy is the uncertainty of the output forecasts. This means that in case of a too small share of the forecasted output is sold as forwards, a long position is left, while for selling too much, a short position is left exposed to market spot prices. To determine the required number of forward contracts needed, the plant operator should estimate the wind farm output pattern in hourly steps for one MW of electricity over a predefined period. Then, this output should be valued by weighting the electricity spot prices with the expected volume which results in an expected revenue per MW for this period. This revenue is divided by the cost of a forward contract for one MWh to determine the right number of forwards needed. An adjustment for wind output that is correlated with the market prices can be made. For a negative correlation, which means that prices are lower when a lot of wind energy is produced, the plant operator should reduce the hedge quantities in order to reduce revenue dearth.¹⁸⁵

As abovementioned, forward contracts usually are not traded for maturities longer than two years. For longer-term hedging, Aydin et al. suggest natural gas swaps since gas contracts are often available for longer maturities.¹⁸⁶ For this hedging strategy, spot gas is bought for settling against fixed gas forward purchase. In order to buy the gas, the spot market revenues of the wind farm are used. The amount of gas is scaled by the expected market heat rate. The latter represents the cost for the amount of gas needed to produce the appropriate quantity of electricity from the gas.¹⁸⁷ Gas is the chosen commodity for hedging as the prices of gas and electricity strongly correlate.¹⁸⁸

As an example, a wind farm with a 100 MW peak power and a 35 % capacity factor for the on-peak period¹⁸⁹ for a specific month is assumed. Furthermore, the average spot price for the on-peak hours in this month is assumed to be 40 € /MWh as well as a forward price of 70 € /MWh. The hedging strategy would then be hedging 20 % which equals 0.2 MW per produced MW.¹⁹⁰ For a gas forward for the same month, the price is presumed to be 4 € /MMBtu which implies a market heat rate of 23.3 MMBtu/MWh. The hedging

strategy for these parameters would be hedging 4.7 MMBtu for each wind MW by using gas swaps.¹⁹¹ A side fact that should be considered is that the market heat rate is not fixed and there might be errors in forecasting. This is based on the fact that gas and electricity prices are not perfectly correlated and the degree of correlation is dependent on the penetration of these commodities as well as the electricity generating technologies in the specific market. Therefore, as Aydin et al. state, a probability-weighted market heat rate should be used.¹⁹²

Aside from derivatives hedging commodity price risk, there are weather and even more specific wind derivatives. Weather derivatives are contracts whose underlying can neither be held nor produced. These contracts are therefore financial tools that can be used to hedge against risk associated with specific weather conditions. They are financial market products and can to some extent also be classified as insurance products. Wind derivatives are traded at the Chicago Mercantile Exchange and OTC. As a major industry, dependent on weather conditions, energy companies are principal investors in the weather market.¹⁹³

Another type of weather derivatives are temperature contracts. The characteristic of the latter is the link to temperature as weather index too. The weather derivative market is growing faster compared to the market for wind derivatives. This is due to the fact that it is very difficult to model wind accurately and to value these contracts appropriately.¹⁹⁴

Wind derivatives are usually standardized products, dependent on the daily average wind speed. This underlying average wind speed is measured over a specified period by a predefined station. Nevertheless, the modeled average wind speeds can only hedge a part of the weather risk as wind sensitivities and power curves vary for different turbines with some plants even requiring a specific duration of a wind speed level. This wind duration factor can be included as underlying weather index in an additional risk hedging strategy to take this requirement into account.¹⁹⁵ The traded volume of contracts on wind conditions is increasing yearly.¹⁹⁶

8.1.3. Insurances

Besides financial instruments like options, there are also insurances which can reduce the risk of volatile revenues. However, common insurance products focus on weather conditions like fluctuating wind and not volatile market prices.

One example is the “lack of wind cover” by Munich Re. This product ensures specified minimum revenues to operators for periods with wind speeds that fall far below expectations. That way, operators meet the financing and operation costs and the return targets. Plant operators are covered

¹⁸³Cf. Deng and Oren (2006), p. 944.

¹⁸⁴Cf. Aydin et al. (2017), p. 5.

¹⁸⁵Cf. Aydin et al. (2017), pp. 5, 20.

¹⁸⁶Cf. Aydin et al. (2017), p. 11.

¹⁸⁷Cf. Aydin et al. (2017), p. 11.

¹⁸⁸Cf. Pschick (2014), p. 40.

¹⁸⁹This usually is during the day time, when demand for electricity is the highest.

¹⁹⁰This ratio is the result of the calculation $\frac{40\text{€ /MWh}}{70\text{€ /MWh}} * 35\%$

¹⁹¹Cf. Aydin et al. (2017), p. 13; This results from multiplying the 20 % hedge ratio with the market heat rate $0.2 * 23.3 = 4.7$.

¹⁹²Cf. Aydin et al. (2017), pp. 21-22.

¹⁹³Cf. Aid (2015), p. 86; Alexandridis and Zapranis (2013), p. 300.

¹⁹⁴Cf. Alexandridis and Zapranis (2013), p. 300.

¹⁹⁵Cf. Alexandridis and Zapranis (2013), pp. 300-301.

¹⁹⁶Cf. Alexandridis and Zapranis (2013), p. 319.

against revenue shortfalls caused by poor wind conditions. In order to set up the specifications of the insurance contract, historical weather data for the location is combined with the turbine specific power-curve and multiplied with the number of turbines to calculate the annual energy yield of the wind farm. If the actual turnover is – as a result of a lack of wind - below the modeled yearly turnover which results from the energy yield multiplied with a fixed price per MWh from a feed-in tariff or an PPA, a certain range will be covered by the insurance.¹⁹⁷ Consequently, this insurance product limits the risk coming from the intermittency of the output, not from volatile market prices. Nevertheless, the variable output is one factor of the hardly predictable revenues.

Another insurance product which covers lost revenues caused by weather in general is the KLIMArisk by the insurance company HDI Global. The contract can be specified to cover shortfalls caused by different weather events like too strong or too weak wind, strong waves or temperature or precipitation extremes. A fixed sum is paid by the insurance company when the abovementioned parametric weather index hits a certain threshold.¹⁹⁸ Even if this kind of insurance is not specifically designed for wind farm operators to hedge against a lack of wind, it can be used for this purpose.

This chapter provides just a brief overview of insurance products on turnover shortfalls due to fluctuating wind. There is probably a variety of comparable products offered on the market such as the weather resource protection product by Allianz that covers electricity generation shortcomings caused by fluctuating wind.¹⁹⁹ As a result, it can be stated, that existing insurances focus on the wind farms energy yield dependent on the wind speed and not on revenues dependent volatile market prices. This might be due to the fact, that at this moment, there are almost predefined prices for electricity in most markets ensured by support schemes.

8.2. Adjustment of the framework for renewables-dominated electricity markets

The increasing share of RES influences the electricity market substantially. It led to lower and more volatile market prices. The reason for this phenomenon are very low marginal cost of electricity from RES on the one hand and the presence of support schemes suppressing market signals on the other hand. Some of the latter still occur by giving incentives for feeding-in energy even if there is already a surplus in supply, characterized by negative prices.²⁰⁰

The financial support schemes introduced in section 3 incentivized investments in RES and contributed significantly to the rapidly growing share of RES in the past. However, they intervene in the market via levering out mechanisms and signals to a certain extent. The expeditious market penetration of renewable energy resulted in low and volatile mar-

ket prices which led to difficulties for conventional electricity producers when it comes to covering expenses. With a high share of intermittent renewables, general market design adjustments are inevitable.²⁰¹ Morch et al. formulate key features of a successful design for RES dominated markets which are shown in figure 14, namely faster and larger markets, smaller products, efficient pricing and a level playing field for all market players.²⁰²

For RES, generation conditions are intermittent and change fast due to weather. These circumstances should be reflected in markets that are RES dominated by allowing for shorter lead times to establish faster markets. Plant operators must provide a generation schedule in advance to the transmission system operator. With shorter lead times, the gate closure for submitting the bids or updating the forecasts moves towards the delivery hour. RES operators would likely be able to self-balance deviations for the variable generation due to more accurate short term forecasts if the gate closure was as close as possible to real time. Consequently, less balancing capacity would be needed.²⁰⁴

Furthermore, larger markets in geographical terms would benefit from smoother and less intermittent output from RES. The phenomenon of reduced overall fluctuations via geographically diversified markets would be taken advantage of. By using a common European grid model, the present infrastructure could probably be exploited best. The enlargement of markets also induces a growing interconnection between regions and cross-border competition.²⁰⁵ Thus, the aggregation of energy markets could result in a more economic integration of intermittent RES.²⁰⁶

The key feature of smaller products refers to the time-frame of contracts. Usually there are one- hour long power contracts traded.²⁰⁷ However, smaller products in terms of shorter timespans makes trading, especially for wind energy, easier. The reason for that is the higher accuracy of output predictions for short time frames. Smaller products also allow adjustments of bids closer to real time. Nonetheless, smaller products should be introduced in combination with larger products as well. This is essential for balancing out liquidity in the markets and implementation cost. A first step towards smaller products that occurred was the implementation of a 15-minute product at the German intraday market.²⁰⁸

Furthermore, pricing in the electricity wholesale market should be efficient. This implies transparent pricing and allowing for volatile prices and spikes that signal scarcity instead of artificially influencing them. This is expedient as transparent prices reflect market signals like the need for

²⁰¹Cf. Morch and Wolfgang (2016), p. 18.

²⁰²Cf. Morch and Wolfgang (2016), p. 21.

²⁰³Own figure based on Morch and Wolfgang (2016), p. 21.

²⁰⁴Cf. Krohn et al. (2009), p. 94; Morch and Wolfgang (2016), p. 22.

²⁰⁵Cf. Morch and Wolfgang (2016), p. 22.

²⁰⁶Cf. Krohn et al. (2009), p. 94.

²⁰⁷Cf. Krohn et al. (2009), p. 96.

²⁰⁸Cf. Morch and Wolfgang (2016), p. 23.

¹⁹⁷Cf. Munich Re (2015).

¹⁹⁸Cf. Gatzert and Kosub (2016), pp. 989-990; HDI Global SE. (n.d.).

¹⁹⁹Cf. Allianz (2015).

²⁰⁰Cf. Morch and Wolfgang (2016), p. 12.

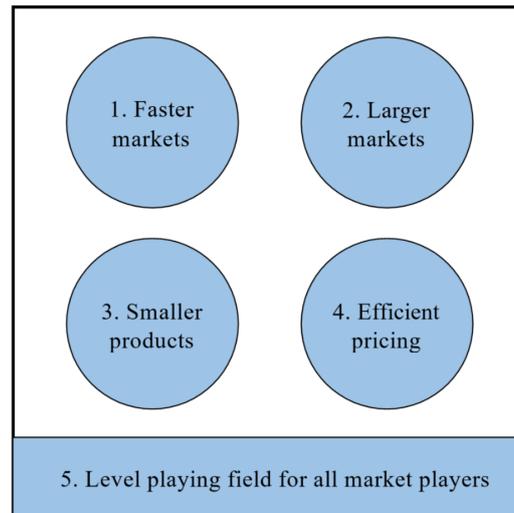


Figure 14: Crucial market characteristics for a successful market integration of RES²⁰³

additional investments. The transparency of prices also includes direct relation to marginal costs of electricity production.²⁰⁹

For all of the mentioned market features, the playing field within the electricity market must be leveled. Firstly, this includes that balancing responsibilities apply to everyone. This becomes more feasible for RES operators with the abovementioned faster markets which allow short gate-closure time. Secondly, the priority feed-in compared to conventional energy existing in some countries must be eliminated to achieve a level playing field. Thirdly, a polluter pays principle should be introduced as the cost of pollution is calculated to low. This results in cheap conventional electricity and artificially uphold competitiveness with respect to RES. And finally, the subsidies for conventional energy sources must be adapted to the support for RES. The high level of subsidies paid to conventional electricity generation must be adjusted analogously to the state aid guidelines for RES.²¹⁰

RES operators would especially benefit from the market features of shorter lead times and smaller products as this would reduce balancing risk and help to stabilize revenues. Nevertheless, for the introduced market framework, the support policy must be adjusted as well. Therefore, the support schemes should stabilize the revenues of RES but also allow for market signals. The subsidies should be allocated using an adequate mechanism like a tender. Furthermore, the paid premiums should be based on produced energy and not on capacity which would bring profits to a generator even without generating. To benefit from a wide portfolio of generating technologies in the long run, supports should be technology-specific. Consequently, further development and cost-reduction for specific technologies can be stimulated. Finally, as well as for the market features, a level playing field

for all market players should be established.²¹¹

8.3. Diversification as an instrument to reduce output volatility

The weather is responsible for the intermittent output of wind power turbines. As the weather deviates at different places at the same time, geographical diversification is an instrument to reduce weather-related output volatility. The basic principle of geographic dispersion is the following. The wind speed and direction is not the same at every place and the correlation of wind speeds decreases for wider distances between wind farms. Small or negative correlations are beneficial in this case as it can be used to smoothen the output fluctuations.²¹² Furthermore, wind forecasts are more precise for large areas like whole countries compared to smaller regions.²¹³ Thus, the overall forecast for a geographically dispersed portfolio of wind farms would be more accurate than the forecast for a single wind farm.

A further step of geographic diversification would be to diversify a portfolio of wind farms over multiple regulatory regimes. Aside from the reduction of the output variability, this also mitigates policy or regulatory risk as the total portfolio revenues are not dependent on the regulatory framework of one single country.²¹⁴

Obviously, the introduced geographic diversification is applicable for portfolios of wind farms only, not single ones. Therefore, this kind of diversification solely constitutes an option for companies with a sufficiently big scale.

9. Conclusion

The objective of this thesis was to analyze the influence a battery storage has on investment risk for wind farm

²⁰⁹Cf. Morch and Wolfgang (2016), p. 23.

²¹⁰Cf. Morch and Wolfgang (2016), pp. 23-24.

²¹¹Cf. Montoro and Corbetta (2016), pp. 25-27.

²¹²Cf. Drake and Hubacek (2007), p. 4001.

²¹³Cf. Holttinen (2005), p. 2057.

²¹⁴Cf. Watts (2011), p. 17.

projects. With progressive market integration of RES, wind farm projects are increasingly exposed to electricity market risks accompanied by uncertainty in future revenues. Existing Literature is extensively addressing the topic of market integration of RES and the forthcoming challenges coming from accelerating exposure to market risks. Similarly, literature on VPPs in general as well as VPPs consisting of wind power and a battery storage exists. However, these studies mostly focus on optimal operation strategies from a risk-neutral point of view. This thesis enlarges the current field of research by pointing out possibilities to reduce investment risk coming from uncertain and volatile revenues. This is necessary to incentivize further investments in RES which is a prerequisite to meet the defined targets on green house gas emissions. The main contribution of this thesis is the analysis of the influence of battery storage on risk coming from revenue volatility in a business case without support schemes and making the leap to cost of capital. Based on the assumption that investors are risk averse and that the WACC declines with lowering risk, the study provides evidence that a battery storage can indeed reduce the overall financial risk.

The results of the study show that deploying a battery in the proposed way benefits both, RES investors and operators. For the latter, stable and plannable revenues as a consequence of energy storage opportunity lead to risk reduction. Investors probably can maintain wind farm projects as low-risk long-term investment options in the future. This furthermore results in low financing cost for operators.

The analysis is based on wind farm and electricity market data from Germany. Results state that for reaching the same NPV, the cost of capital for a VPP has to be reduced compared to an usual wind park investment. The scale of the required WACC reduction reaches from less than 0.5 up to more than 4.0 percentage points depending on input data assumptions about size and price of the battery, the project's timeframe and reference rate for cost of capital. Since the operation of an VPP like the constellation introduced in the case study will be most reasonable post 2020, the assumptions made represent the main limitation of the model. Therefore, assumptions on battery prices are made based on predictions. Moreover, the technological characteristics of the chosen battery such as partial load limit, operating cost or cycle stability could be included more detailed to map the operation more realistically. In addition, it is questionable how required returns from RES investments and consequently cost of capital will evolve during times when no certainty by state-aid is provided. As e.g. Doetsch et al. point out, the determination of the actual level of cost of capital by investors and banks is to a certain extent subjective and influenced by qualitative measures like political stability as well.²¹⁵ Furthermore, the volatility and height of the prices used for the study might not be representative for the post 2020 era. The reason for this is the increasing share of intermittent energy sources whose low marginal cost and stochastic generation cause not only

lower but also more volatile market prices.²¹⁶

An extension of the model would be to add generality by formulating it in a stochastic way. The objective function could then be formulated not in order to maximize the revenue and assess the influence on the risk afterwards but to optimize the impact on the investment risk. This kind of modelling was implemented in literature for general VPPs in terms of VaR maximization. Furthermore, the battery size could be included as a variable to identify optimum battery parameters. However, this would add further complexity. Thus, a deterministic model was chosen for this thesis to increase clarity of the results which is likely to result in wider acceptance for practical use.²¹⁷

Circumstances for grid integrated storage deployment will probably change in the future. This is due to the abovementioned lower electricity prices and increasing price volatility as well as the ceasing of state-aid. However, especially the falling prices and ongoing development of batteries will probably be game-changing for stationary electricity storage. With a larger scale of installed electricity storage capacity and lower cost of storing energy, the trading of electricity changes as it becomes an affordably storable commodity. Assuming sufficient installed storage capacity, a wind farm can decouple its supply profile from the actual electricity production curve.²¹⁸ With more affordable batteries, storages can also be used as multi-application devices aside from the VPP function. Additional revenues may be generated via arbitrage trading, providing services like reserve supply or frequency stabilization, or by offering balancing capacity. Consequently, revenues can be generated by different tasks to make battery systems more valuable. A further business case for using storages was introduced by Pinson et al., namely offering electricity storage capacity as a service in the form of an independent electricity market entity. This would enable RES operators to rent capacity for reducing tied-up capital and determining the required storage size dynamically in order to avoid under-utilization in low wind times.²¹⁹

Furthermore, the deployment of a second-use storage by Daimler and The Mobility House provides strong evidence that the application in stationary storages is an economically viable lifetime extension for EV batteries.²²⁰ With an increasing share of EVs, more used batteries will be available in the market. This provides inexpensive storage capacity on the one hand, and improves the batteries' eco-balance on the other hand. The mentioned technological evolution and further innovation on smart grids makes the issue of storing electricity in large scales even more tangible. This is however crucial to make RES dispatchable and therefore competitive in the market. Combined with revenue-stabilizing instruments, RES are likely to remain attractive investments in the future.

²¹⁵Cf. Doetsch et al. (2011), p. 3.

²¹⁶Cf. Ketterer (2012), p. 35.

²¹⁷Cf. Cummins et al. (1999), p. 424.

²¹⁸Cf. Rugolo and Aziz (2012), p. 7159.

²¹⁹Cf. Pinson et al. (2009), p. 7.

²²⁰Cf. The Mobility House (2016).

Dealing with this topic becomes increasingly important to reach the defined target shares of RES in the market. This is indispensable for cutting the carbon emissions, counteracting climate change and contributing to a sustainable future.

References

- Aïd, R. *Electricity derivatives*. Springer, 2015.
- Alexandridis, A. and Zapranis, A. Wind derivatives: Modeling and pricing. *Computational Economics*, 41(3):299–326, 2013.
- Allianz. Sectors - Renewable Energy, 2015. URL <https://www.agcs.allianz.com/sectors/energy-renewables/>. 13/04/2017.
- Aydin, C. O., Graves, F. C., and Villadsen, B. Managing Price Risk for Merchant Renewable Investments: Role of Market Interactions and Dynamics on Effective Hedging Strategies, The Brattle Group, 2017.
- Baker & McKenzie. The rise of corporate PPAs - A new driver for renewables, 2015. URL <https://www.bakermckenzie.com/en/insight/publications/2015/12/the-rise-of-corporate-ppas>. 10/04/2017.
- Barton, J. P., Infield, D. G., et al. Energy storage and its use with intermittent renewable energy. *IEEE transactions on energy conversion*, 19(2):441–448, 2004.
- Becker, M. Spiegel Online: EU-Kommission will Energiemarkt umkrempeln, 2016. URL <http://www.spiegel.de/wirtschaft/soziales/winterpaket-eu-kommission-will-mehr-energie-effizienz-a-1123782.html>. 22/02/2017.
- Better World Solutions. US Energy Storage Update, 2015. URL <https://www.betterworldsolutions.eu/us-energy-storage-update/>. 20/03/2017.
- Binda Zane, E., Brückmann, R., Bauknecht, D., Jirous, F., Piria, R., Trennepohl, N., Bracker, J., Frank, R., and Herling, J. Integration of electricity from renewables to the electricity grid and to the electricity market – RES-INTEGRATION - Final Report, Berlin, 2012. URL <http://www.eclareon.com/de/projects/res-integration-final-report>.
- Bishop, S. and Officer, R. R. Review of Debt Risk Premium and Market Risk Premium, Melbourne et al.: Value Adviser Associates, 2013.
- Bloomberg New Energy Finance. Corporate Renewable Energy Procurement Monthly - December 2016, 2016. URL <https://about.bnef.com/blog/corporate-renewable-energy-procurement-monthly-december/>. 20/04/2017.
- Bundesministerium für Wirtschaft und Energie. Die erneuerbaren Energien - Fit für den Strommarkt. Fit für die Zukunft - Alle wichtigen Fakten zum neuen EEG 2017. Berlin, from:, 2016. URL <https://www.bmwi.de/BMWi/Redaktion/PDF/Publikationen/fit-fuer-den-strommarkt,property=pdf,bereich=bmwi2012,sprache=de,rwb=true.pdf>.
- Cleijne, H. and Ruijgrok, W. Modelling Risks of Renewable Energy Investments, Green-X, from:, 2004. URL [http://www.green-x.at/downloads/WP2%20-%20Modelling%20risks%20of%20renewable%20energy%20investments%20\(Green-X\).pdf](http://www.green-x.at/downloads/WP2%20-%20Modelling%20risks%20of%20renewable%20energy%20investments%20(Green-X).pdf).
- Conejo, A. J., Plazas, M. A., Espinola, R., and Molina, A. B. Day-ahead electricity price forecasting using the wavelet transform and arima models. *IEEE transactions on power systems*, 20(2):1035–1042, 2005.
- Costa, L. M., Bourry, F., Juban, J., and Kariniotakis, G. Management of energy storage coordinated with wind power under electricity market conditions. In *Probabilistic Methods Applied to Power Systems, 2008. PMAPS'08. Proceedings of the 10th International Conference on*, pages 1–8. IEEE, 2008.
- Cummins, J. D., Grace, M. F., and Phillips, R. D. Regulatory solvency prediction in property-liability insurance: Risk-based capital, audit ratios, and cash flow simulation. *Journal of Risk and Insurance*, pages 417–458, 1999.
- Damodaran, A. *Damodaran on Valuation - Security Analysis for Investment and Corporate Finance*. Wiley, Hoboken, New Jersey, 2006.
- D'Aprile, P., Newman, J., and Pinner, D. The new economics of energy storage, from:, 2016. URL <http://www.mckinsey.com/business-functions/sustainability-and-resource-productivity/our-insights/the-new-economics-of-energy-storage>. 09/02/2017.
- Deign, J. Wind Energy Update - Investors look to combine PV and wind assets, from:, 2014. URL <http://analysis.windenergyupdate.com/operations-maintenance/investors-look-combine-pv-and-wind-assets>. 29/04/2017.
- Deng, S.-J. and Oren, S. S. Electricity derivatives and risk management. *Energy*, 31(6-7):940–953, 2006.
- DLA Piper. The year of PPAs and the corporate green agenda, from:, 2016. URL <https://www.dlapiper.com/en/uk/insights/publications/2016/06/renewable-energy-global-paper/what-are-corporate-power-purchase-agreements-ppa/>. 11/04/2017.
- Doetsch, C., Kanngießer, A., Wolf, D., Schinz, S., Sperling, M., Frey, H., and Kamga, A. K. Netzintegrierte Stromspeicher zur Integration fluktuierender Energie – Technische Anforderungen, ökonomischer Nutzen, reale Einsatzszenarien - Abschlussbericht. Fraunhofer UMSICHT, TU Darmstadt, EnBW, from:, 2011. URL http://publica.fraunhofer.de/eprints/urn_nbn_de_0011-n-1901398.pdf.
- Drake, B. and Hubacek, K. What to expect from a greater geographic dispersion of wind farms?—a risk portfolio approach. *Energy Policy*, 35(8):3999–4008, 2007.
- Dunn, B., Kamath, H., and Tarascon, J.-M. Electrical energy storage for the grid: a battery of choices. *Science*, 334(6058):928–935, 2011.
- Ekren, B. Y. and Ekren, O. Simulation based size optimization of a pv/wind hybrid energy conversion system with battery storage under various load and auxiliary energy conditions. *Applied Energy*, 86(9):1387–1394, 2009.
- Energiespeicher - Forschungsinitiative der Bundesregierung. Vanadium-Redox-Flow - Großspeicher in Betrieb genommen, from:, 2015. URL http://forschung-energiespeicher.info/aktuelles/aktuelles-einzelansicht/2/Grossspeicher_in_Betrieb_genommen/. 12/01/2017.
- European Climate Foundation. Roadmap 2050 - A practical guide to a prosperous, low-carbon europe - Technical Analysis executive summary, from:, 2010. URL http://www.roadmap2050.eu/attachments/files/Volume1_ExecutiveSummary.pdf.
- European Commission. European Commission guidance for the design of renewables support schemes. Brussels, from:, 2013. URL <https://ec.europa.eu/energy/en/topics/renewable-energy/support-schemes>.
- European Commission. Guidelines on State aid for environmental protection and energy 2014-2020. Communication from the Commission (2014/C 200/01), in: Official Journal of the European Union C 200(28.6.2014), pp. 1-55, 2014.
- European Commission. EU Reference Scenario 2016 - Energy, transport and GHG emissions trends to 2050. Luxembourg, from:, 2016. URL https://ec.europa.eu/energy/sites/ener/files/documents/ref2016_report_final-web.pdf.
- European Commission. European Commission - Energy - Support Schemes, from:, 2017. URL <https://ec.europa.eu/energy/en/topics/renewable-energy/support-schemes23/04/2017>.
- European Commission. Climate action - 2030 climate & energy framework, from:, n.d. URL http://ec.europa.eu/clima/policies/strategies/2030/index_en.htm. 02/11/2016.
- Eyer, J. and Corey, G. Energy Storage for the Electricity Grid: Benefits and Market Potential Assessment Guide. Albuquerque: Sandia National Laboratories, from:, 2010. URL www.sandia.gov/ess/publications/SAND2010-0815.pdf.
- Fama, E. F. and French, K. R. The cross-section of expected stock returns. *the Journal of Finance*, 47(2):427–465, 1992.
- Fama, E. F. and French, K. R. The capital asset pricing model: Theory and evidence. *Journal of economic perspectives*, 18(3):25–46, 2004.
- Frontline Systems, Inc. Excel Solver - Algorithms and methods used, from:, 2017. URL <http://www.solver.com/excel-solver-algorithms-and-methods-used>. 08/03/2017.
- Gatzert, N. and Kosub, T. Risks and risk management of renewable energy projects: The case of onshore and offshore wind parks. *Renewable and Sustainable Energy Reviews*, 60:982–998, 2016.
- Gersema, G. and Wozabal, D. Risk Optimal Pooling of Intermittent Renewable Energy Resource, in *Journal of Banking and Finance*, forthcoming, from:, 2016. URL <https://ssrn.com/abstract=2784736>.
- Gillhuber, A. Hybrid-Stromspeicher - Eine Woche Strom für 40 Häuser?, from:, 2014. URL <http://www.elektroniknet.de/elektronik/power/eine-woche-strom-fuer-40-haeuser-111844.html>. 20/03/2017.
- Gönsch, J. and Hassler, M. Sell or store? an adp approach to marketing renewable energy. *OR spectrum*, 38(3):633–660, 2016.
- Green Rhino Energy. Green Rhino Energy - Energy Price Risk, from:, 2013. URL http://www.greenrhinoenergy.com/finance/renewable/energy_price_risk.php. 18/04/2017.
- Günter, N. and Marinopoulos, A. Energy storage for grid services and applications: Classification, market review, metrics, and methodology for evaluation of deployment cases. *Journal of Energy Storage*, 8:226–234, 2016.
- Hanania, J., Stenhouse, K., and Donev, J. Energy Education - Intermittent electricity, from:, n.d. URL <http://energyeducation.ca/encyclo>

- dia/Intermittent_electricity. 01/04/2017.
- HDI Global SE. KLIMARisk - Die Wetterrisikoversicherung, from:, n.d. URL <https://www.hdi.global/de/de/versicherungen/sachversicherungen/wetterrisikoversicherung>. 13/04/2017.
- Held, A., Ragwitz, M., Gephart, M., De Visser, E., and Klessmann, C. Design features of support schemes for renewable electricity. *A report within the European project "Cooperation between EU MS under the Renewable Energy Directive and interaction with support schemes"*. Ecofys Netherlands, Utrecht, 2014.
- Hippert, H. S., Pedreira, C. E., and Souza, R. C. Neural networks for short-term load forecasting: A review and evaluation. *IEEE Transactions on power systems*, 16(1):44–55, 2001.
- Hirth, L. The optimal share of variable renewables: How the variability of wind and solar power affects their welfare-optimal deployment. *Energy Journal*, 36(1):149–184, 2015.
- Holttinen, H. Optimal electricity market for wind power. *Energy Policy*, 33(16):2052–2063, 2005.
- Investopedia. Investopedia - Swap, from:, 2017. URL <http://www.investopedia.com/terms/s/swap.asp>20/04/2017. 20/04/2017.
- IWR. Der IWR-Windertragsindex für Regionen 5-jähriger Index 2013, from:, 2014. URL http://www.iwr.de/wind/wind/windindex/index13_5jahre.htm. 30/04/2017.
- IWR. Der IWR-Windertragsindex für Regionen 5-jähriger Index 2014, from:, 2015. URL http://www.iwr.de/wind/wind/windindex/index14_5jahre.htm. 30/04/2017.
- IWR. Der IWR-Windertragsindex für Regionen 5-jähriger Index 2015, from:, 2016. URL http://www.iwr.de/wind/wind/windindex/index15_5jahre.htm. 30/04/2017.
- Jorion, P. *Value at Risk - The New Benchmark for Managing Financial Risk (3rd edition)*. McGraw-Hill, 2006.
- Kempton, W. and Tomić, J. Vehicle-to-grid power implementation: From stabilizing the grid to supporting large-scale renewable energy. *Journal of power sources*, 144(1):280–294, 2005.
- Ketterer, J. C. The Impact of Wind Power Generation on the Electricity Price in Germany - Ifo Working Paper No. 143. Ifo Institute - Leibniz Institute for Economic Research, Munich, from:, 2012. URL <https://www.cesifo-group.de/portal/page/portal/CBC3441402262FCAE04400144FAFBA7C>.
- Klessmann, C., Nabe, C., and Burges, K. Pros and cons of exposing renewables to electricity market risks—a comparison of the market integration approaches in germany, spain, and the uk. *Energy Policy*, 36(10):3646–3661, 2008.
- Krohn, S., Morthorst, P.-E., and Awerbuch, S. The economics of wind energy. *The European Wind Energy Association (EWEA)*, 2009.
- Lassila, J., Haakana, J., Tikka, V., and Partanen, J. Methodology to analyze the economic effects of electric cars as energy storages. *IEEE Transactions on smart grid*, 3(1):506–516, 2012.
- Leuthner, S. Übersicht zu lithium-ionen-batterien. In *Handbuch Lithium-Ionen-Batterien*, pages 13–19. Springer, 2013.
- Liu, Y., Jiang, C., Shen, J., Hu, J., and Luo, Y. Coordination of hydro units with wind power generation based on raroc. *Renewable Energy*, 80:783–792, 2015.
- Modigliani, F. and Miller, M. H. The cost of capital, corporation finance and the theory of investment. *The American economic review*, 48(3):261–297, 1958.
- Moghaddam, I. G., Nick, M., Fallahi, F., Sanei, M., and Mortazavi, S. Risk-averse profit-based optimal operation strategy of a combined wind farm-cascade hydro system in an electricity market. *Renewable energy*, 55: 252–259, 2013.
- Montoro, D. F. and Corbetta, G. Making transition work. WindEurope, 2016.
- Morch, A. and Wolfgang, O. Post-2020 framework for a liberalised electricity market with a large share of renewable energy sources. *Final Publication of the Market4Res project*, 2016.
- Munich Re. Lack of Wind Cover, from:, 2015. URL <https://www.munichre.com/weatherandcommodity/en/group/index.html>. 13/04/2017.
- Nicola, S.-L. RES legal - Legal sources on renewable energy - Premium tariff (Market Premium), from:, 2017a. URL <http://www.res-legal.eu/search-by-country/germany/single/s/res-e/t/promotion/aid/premium-tariff-i-market-premium/lastp/135/>. 29/04/2017.
- Nicola, S.-L. RES legal - Legal sources on renewable energy - Promotion in Germany, from:, 2017b. URL <http://www.res-legal.eu/search-by-country/germany/tools-list/c/germany/s/res-e/t/promotion/sum/136/lpid/135/>. 29/04/2017.
- Noothout, P., de Jager, D., Tesnière, L., van Rooijen, S., Karypidis, N., Brückmann, R., Jirouš, F., Breitschopf, B., Angelopoulos, D., Doukas, H., et al. The impact of risks in renewable energy investments and the role of smart policies. *DiaCore report*, 2016.
- Oberthür, S. and Ott, H. E. *The Kyoto Protocol: international climate policy for the 21st century*. Springer Science & Business Media, 1999.
- Overton, T. W. SDG&E Unveils World's Largest Li-Ion Storage Battery, from:, 2017. URL <http://www.powermag.com/sdge-unveils-worlds-largest-li-ion-storage-battery/?printmode=0>. 19/03/2017.
- Penndorf, S. Renewable Energy Power Purchase Agreements, from:, 2016. URL <https://3degreesinc.com/ppas-power-purchase-agreements/>. 12/04/2017.
- Pinson, P., Papaefthymiou, G., Klockl, B., and Verboomen, J. Dynamic sizing of energy storage for hedging wind power forecast uncertainty. In *Power & Energy Society General Meeting, 2009. PES'09. IEEE*, pages 1–8. IEEE, 2009.
- Poole, I. Lead Acid Battery Tutorial, from:, 2017a. URL <http://www.radio-electronics.com/info/power-management/battery-technology/lead-acid-battery-tutorial.php>. 18/03/2017.
- Poole, I. Lithium Ion Battery Advantages & Disadvantages, from:, 2017b. URL <http://www.radio-electronics.com/info/power-management/battery-technology/lithium-ion-battery-advantages-disadvantages.php>. 18/03/2017.
- Pschick, A. W. Hedgingstrategien im Stromgroßhandel: Preis- und Kreditrisiken sicher im Griff, Disserta Verlag, Hamburg, 2014.
- Rathmann, M., de Jager, D., De Lovinofosse, I. d., Breitschopf, B., Burgers, J., and Weöres, B. Towards triple-a policies: More renewable energy at lower cost. *Re-Shaping D16 Report*, 2011.
- RE100. RE100 - Companies, from:, n.d. URL <http://there100.org/companies>. 11/04/2017.
- Rosenbaum, J. and Pearl, J. *Investment banking: valuation, leveraged buy-outs, and mergers and acquisitions*. John Wiley & Sons, 2013.
- Rottmann, H., Lackses, R., Siepermann, M., Auer, B. R., and Lübbecke, M. *Gabler Wirtschaftslexikon - Simulation*. (S. G. Verlag, Editor), from:, n.d. URL <http://wirtschaftslexikon.gabler.de/Archiv/55029/simulation-v13.html>. 06/05/2017.
- Rugolo, J. and Aziz, M. J. Electricity storage for intermittent renewable sources. *Energy & Environmental Science*, 5(5):7151–7160, 2012.
- Salles, M., Aziz, M. J., and Hogan, W. W. Potential arbitrage revenue of energy storage systems in pjm during 2014. In *Power and Energy Society General Meeting (PESGM), 2016*, pages 1–5. IEEE, 2016.
- Sioshansi, R. Increasing the value of wind with energy storage. *The Energy Journal*, pages 1–29, 2011.
- Staffell, I. and Rustomji, M. Maximising the value of electricity storage. *Journal of Energy Storage*, 8:212–225, 2016.
- Statista. Wirkungsgrade verschiedener Stromspeicher im Jahr 2012, from:, n.d. URL <https://de.statista.com/statistik/daten/studie/156269/umfrage/wirkungsgrade-von-ausgewaehlten-stromspeichern/>. 21/03/2017.
- Structured Data, LLC. RiskAMP is a full-featured Monte Carlo Simulation Engine for Microsoft Excel, from:, 2017. URL <https://www.riskamp.com/>. 09/03/2017.
- Structured Data LLC. What is Monte Carlo Simulation?, from:, n.d. URL <https://www.riskamp.com/files/RiskAMP%20-%20MonteCarlo%20Simulation.pdf>. 06/05/2017.
- The Mobility House. World's Largest 2nd-use Battery Storage Is Starting Up, from:, 2016. URL <http://mobilityhouse.com/en/worlds-largest-2nd-use-battery-storage-is-starting-up/>. 05/01/2017.
- United Nations. United Nations Framework Convention on Climate Change - A Summary of the Kyoto Protocol, from:, 2014. URL http://unfccc.int/kyoto_protocol/background/items/2879.php. 01/04/2017.
- Vasirani, M., Kota, R., Cavalcante, R. L., Ossowski, S., and Jennings, N. R. An agent-based approach to virtual power plants of wind power generators and electric vehicles. *IEEE Transactions on Smart Grid*, 4(3):1314–1322, 2013.
- von Hirschhausen, C., Holz, F., Gerbaulet, C., and Lorenz, C. European energy sector: large investments required for sustainability and supply security. *DIW Economic Bulletin*, 4(7):31–36, 2014.
- Wall Street Oasis. Wall Street Oasis - Discounted Cash Flow Definition, from:,

- n.d. URL <https://www.wallstreetoasis.com/finance-dictionary/what-is-a-discounted-cash-flow-DCF>. 22/04/2017.
- Wallasch, A.-K., Lüers, S., and Rehfeldt, K. Kostensituation der Windenergie an Land in Deutschland - Update. Deutsche WindGuard GmbH, Varel, from:, 2015. URL <http://www.windguard.de/service/knowledge-center/veroeffentlichungen.html>. 29/04/2017.
- Watts, C. Managing the risk in renewable energy - A report from the Economist Intelligence Unit Sponsored by Swiss Re. The Economist, 2011.
- Welch, I. Corporate Finance - An Introduction. Bosten et al.: Pearson - Prentice Hall, 2009.
- Wind-Turbine. Weiterbetrieb von Windkraftanlagen – die Ausgangssituation in Deutschland, from:, 2016a. URL <https://wind-turbine.com/magazin/ratgeber/weiterbetrieb/46745/weiterbetrieb-von-windkraftanlagen-die-ausgangssituation-in-deutschland.html>. 02/11/2016.
- Wind-Turbine. Weshalb sich der Weiterbetrieb von Windkraftanlagen lohnt, from:, 2016b. URL <https://wind-turbine.com/magazin/ratgeber/weiterbetrieb/46990/weshalb-sich-der-weiterbetrieb-von-windkraftanlagen-lohnt.html>. 15/02/2017.
- WindEurope. Creating a Business Case for Wind after 2020, from:, 2017. URL <https://windeurope.org/policy/position-papers/>. 25/03/2017.
- Wozabal, D., Graf, C., and Hirschmann, D. The effect of intermittent renewables on the electricity price variance. *OR spectrum*, 38(3):687–709, 2016.
- Yunicos. NOTREES - Wind and Energy Storage Facility, from:, 2016. URL <https://www.yunicos.com/de/case-studies/notrees-usa/>. 20/03/2017.



Konzeptualisierung und Operationalisierung von Patientenzufriedenheit mit Hausärztinnen und Hausärzten

Bernhard Guetz

Alpen-Adria-Universität Klagenfurt

Abstract

In der vorliegenden Arbeit werden verschiedene einflussnehmende Faktoren der Patientenzufriedenheit definiert und unterschiedliche Methoden der Patientenzufriedenheitserhebung dargelegt, wobei der Themenbereich durch umfassende Literatur- und Quellenanalyse übersichtlich aufbereitet wird.

Die Darstellung von Patientenzufriedenheit als eigenständiges Zufriedenheitskonstrukt, neben individueller Zufriedenheit und Kundenzufriedenheit, begründet sich dadurch, dass das Verhältnis zwischen Ärztinnen / Ärzten und Patientinnen / Patienten nicht mit einem reinen Konsumerlebnis und nur am Rande mit einem Dienstleistungsprozedere vergleichbar ist.

Im Bereich der Faktorbestimmung der Patientenzufriedenheit sind häufig Modelle mit drei Überkategorien zu finden, wobei in der Regel nur eine dieser Kategorien durch die Hausärztin / den Hausarzt beeinflussbare Faktoren oder Faktorenbündel enthält. Die Definition der beeinflussbaren Patientenzufriedenheitskategorien kann je nach gewähltem Forschungsschwerpunkt sehr unterschiedlich ausfallen. Häufig genannte und untersuchte Zufriedenheitsdimensionen in dieser Hinsicht sind die empfundenen Methoden- und Fachkenntnisse der behandelnden Ärztin / des behandelnden Arztes, die Freundlichkeit des Praxispersonals, die verbale Kommunikation und andere zwischenmenschliche Aspekte. Im Bereich der Patientenzufriedenheitsmessung kann grundsätzlich zwischen Methoden der Befragung, Methoden der Beobachtung, experimentellen Forschungsdesigns und dem Einsatz von Onlinere Ressourcen zur Erhebung der Patientenzufriedenheit unterschieden werden. Obwohl nicht jede Erhebungsmethode in jedem Kontext eingesetzt werden kann bzw. sollte, wird durch den Einsatz bestimmter Methoden die Untersuchung expliziter Forschungsschwerpunkte begünstigt.

Die gesammelten Ergebnisse zeigen, dass eine Vielzahl unterschiedlicher Erhebungsmethoden im Bereich der Patientenzufriedenheitsmessung mit Hausärztinnen und -ärzten vorteilhaft eingesetzt werden kann, wobei die einflussnehmenden Faktoren, wenn auch unterschiedlich stark ausgeprägt, häufig fachliche und zwischenmenschliche Merkmale beinhalten.

Keywords: Patientenzufriedenheit, Kundenzufriedenheit, Zufriedenheitsmessung, Arzt-Patienten-Beziehung

1. Einleitung und Problemstellung

Die Bedeutung der empfundenen Patientenzufriedenheit hat in jüngster Vergangenheit aus Sicht der Ärztinnen und Ärzte sowie Patientinnen und Patienten stark zugenommen. Diese Entwicklung lässt sich grundsätzlich auf zwei gesellschaftliche Phänomene zurückführen.¹ Einerseits hat sich die Rolle von Ärztinnen und Ärzten wie auch Patientinnen und Patienten seit Mitte des zwanzigsten Jahrhunderts stark verändert,² andererseits sieht sich auch das moderne Gesundheitswesen mit erheblicher wirtschaftlicher Belastung kon-

frontiert.³

Darüber hinaus haben insbesondere die Entwicklungen des Onlinesektors zu einer Reihe von Veränderungen bzgl. des Patientenverhaltens geführt. So ist es nicht nur üblich geworden, bei auftretenden Krankheitssymptomen online nach Diagnosen zu suchen, auch bei der Wahl der Hausärztin / des Hausarztes vertrauen immer mehr Patientinnen und Patienten auf Ratschläge und Meinungen aus sozialen Medien und Ärztebewertungsportalen.⁴ Diese Tendenz hat auch erheblichen Einfluss auf die Positionierung von Ärztinnen und Ärzten, da den Aussagen von Konsumentinnen und Konsu-

¹Vgl. Dampz et al. (2012), S. 8.

²Vgl. Kaba und Sooriakumaran (2007), S. 60.

³Vgl. Dampz et al. (2012), S. 8.

⁴Vgl. Paul et al. (2013), S. 1.

menten häufig größeres Vertrauen zugesprochen wird als institutionellen Veröffentlichungen oder Werbebotschaften. In diesem Zusammenhang ist es für den wirtschaftlichen Erfolg unumgänglich, Patientinnen und Patienten langfristig zufriedenzustellen.⁵

Zufriedene Patientinnen und Patienten wiederholen in der Regel ihre positiven Erfahrungen und legen diese auch anderen nahe. Infolgedessen hat auch die Erhebung der Patientenzufriedenheit an Bedeutung gewonnen, da zufriedene Patientinnen und Patienten eine Grundvoraussetzung für dauerhaftes ökonomisches Durchsetzungsvermögen von Arztpraxen darstellen.⁶ Vor diesem Hintergrund haben Patientenzufriedenheitsstudien in den vergangenen 20 Jahren deutlich an Einfluss gewonnen und sind damit zur zentralen Quelle für die effektive Entwicklung qualitätsverbessernder Maßnahmen im Gesundheitswesen avanciert.⁷

Angesichts der Fülle an Literatur und Untersuchungen hinsichtlich Patientenzufriedenheit wurden verschiedenste Modelle und Messverfahren entwickelt, welche nicht immer zu vergleichbaren Ergebnissen führen. Überdies werden die einzelnen Zufriedenheitsdimensionen häufig unterschiedlich definiert. Dieser Umstand erschwert unter anderem die Vergleichbarkeit und die Messung der Qualitätskriterien sowie die Darstellung zeitgemäßer Modelle und Konzepte im Bereich der Patientenzufriedenheit.

1.1. Zielsetzung

Die Arbeit verfolgt das Ziel, den prozessualen Ablauf der Untersuchung von Patientenzufriedenheit mit Hausärztinnen und -ärzten darzustellen. Diese Charakterisierung soll durch die Beschreibung von logisch aufeinander aufbauenden Prozessschritten erreicht werden. Hierzu sollen im ersten Schritt unterschiedliche Einflussfaktoren auf die subjektiv empfundene Patientenzufriedenheit mit Hausärztinnen und -ärzten definiert und kategorisiert werden. Danach soll untersucht werden, aus welchen Kategorien sich die identifizierten Faktorbündel zusammensetzen. Um diesen Anspruch zu erfüllen, soll dargestellt werden, welche Skalen in der Vergangenheit im Bereich der Patientenzufriedenheitsmessung verwendet wurden und welche Faktorkategorien diese messen bzw. aus welchen Zufriedenheitsdimensionen sich diese zusammensetzen. Ferner soll beispielhaft dargestellt werden, welche Messverfahren in der Vergangenheit angewandt wurden um die Ausprägung der jeweiligen Zufriedenheitsdimensionen zu bestimmen.

Vordergründig soll demnach dargelegt werden, wie Patientenzufriedenheit in der Fachliteratur kategorisiert wird und welche Skalen entwickelt wurden, um Patientenzufriedenheit zu messen. Des Weiteren soll elaboriert werden, welche Methoden sich zur Messung der Patientenzufriedenheit mit Hausärztinnen und -ärzten eignen, wie jene Methoden angewendet werden bzw. in der Vergangenheit angewandt

wurden und welche Faktoren in dieser Hinsicht die subjektiv empfundene Zufriedenheit aus Patientinnen- und Patientensicht beeinflussen.

Ziel der Arbeit ist die Herstellung einer Grundlage zur Durchführung eigener praktischer oder theoretischer Patientenzufriedenheitserhebungen, mit klarem Fokus auf die Erhebungsphase.

1.2. Aufbau der Arbeit und Darstellung der Vorgehensweise

Zu Beginn der Arbeit wird auf begriffliche und konzeptionelle Grundlagen in der Sphäre der Patientenzufriedenheit eingegangen. So werden die Begriffe Zufriedenheit, Kundenzufriedenheit und Patientenzufriedenheit definiert und hinsichtlich Gemeinsamkeiten und Unterschiede aufbereitet. Danach wird näher auf Konzepte und Modelle eingegangen, welche sich mit der Entwicklung von Kunden- und Patientenzufriedenheit auseinandersetzen. Um den Behandlungsprozess von anderen Dienstleistungsverhältnissen abzugrenzen, bilden den Kapitelabschluss Überlegungen hinsichtlich der Besonderheiten innerhalb der Arzt-Patienten-Beziehung.

Im nächsten Kapitel, Einflussfaktoren auf die Patientenzufriedenheit, werden unterschiedliche Ansätze der Kategorisierung von Patientenzufriedenheit abgebildet. Diese Vorgehensweise wurde gewählt, da auch bei der Durchführung eigener Patientenzufriedenheitserhebungen der erste logische Schritt darin besteht, vorhandene Quellen, welche sich mit dem jeweiligen Themenbereich auseinandersetzen, hinsichtlich Inhalt, Aktualität und Passgenauigkeit zu untersuchen. Zu diesem Zweck werden drei verschiedene Modelle zur Grobkategorisierung von Patientenzufriedenheit dargestellt und analysiert, aus welchen Faktoren sich die einzelnen Zufriedenheitskategorien zusammensetzen. Kapitelabschließend wird präzisiert, welche, durch die behandelnde Ärztin / den behandelnden Arzt beeinflussbaren, Patientenzufriedenheitsfaktoren häufig genannt und untersucht werden und welche Methoden angewandt wurden, um diese zu erforschen.

Danach werden verschiedene Skalen zur Messung von Patientenzufriedenheit untersucht und vorgestellt, wobei hier internationale bzw. multinationale und deutschsprachige Skalen zur Messung der Patientenzufriedenheit voneinander abgegrenzt sind. Wie zu Beginn des Kapitels beschrieben, entstehen durch die Untersuchung vorhandener Messskalen zwei vorteilhafte Ausprägungen. Einerseits kann die vorhandene Skala unter Umständen verwendet werden, um die eigene Erhebung durchzuführen, andererseits kann die Orientierung an vorhandenen Skalen dabei helfen, eigene Erhebungsinstrumente zu entwickeln. Darüber hinaus wird durch die Untersuchung vorhandener Skalen zur Messung der Patientenzufriedenheit deutlich, welche Zufriedenheitsdimensionen diese messen bzw. welche Faktorkategorien diese bündeln. Insgesamt werden fünf internationale bzw. multinationale und zwei deutschsprachige Skalen zur Messung der Patientenzufriedenheit vorgestellt, wobei sich die Kategorisierung ausschließlich auf die Ursprungsentwicklungssprache bezieht. Ausgewählte internationale bzw. multinationale Skalen sind durchaus auch in einer deutschen

⁵Vgl. Solomon (2013), S. 435 ff.

⁶Vgl. Geierhos et al. (2015), S. 3 f.

⁷Vgl. Al-Abri und Al-Balushi (2014), S. 3.

Version erhältlich und eine der deutschsprachigen Skalen wurde auch zur Erhebung der Patientenzufriedenheit in anderen Sprachräumen verwendet. Die Abgrenzung der beiden Kategorien wurde gewählt, da durch die Such- und Auswahlmethode internationale Skalen begünstigt wurden und dadurch die in deutscher Sprache entwickelten Skalen innerhalb der Arbeit unterrepräsentiert gewesen wären. Im Resümee werden die identifizierten Faktorkategorien den Dienstleistungsdimensionen, Potentialdimensionen, Prozessdimensionen oder Ergebnisdimensionen zugeordnet und die unterschiedlichen internationalen und deutschsprachigen Skalen übersichtlich gegenübergestellt.

Abschließend wird näher auf ausgewählte Methoden zur Erhebung der subjektiv empfundenen Patientenzufriedenheit eingegangen. Im Bereich der qualitativen Befragungsmethoden werden Tiefeninterview, Gruppendiskussion und Critical Incident Technique sowie Beschwerde- und Vorschlagssysteme als Instrumente qualitativer Patientenzufriedenheitserhebungen dargestellt. Den Bereich der quantitativen Befragungsmethoden bilden innerhalb der Arbeit die standardisierten Erhebungsformen der mündlichen, schriftlichen und telefonischen Befragung sowie die Onlinebefragung. Ferner wird auch näher auf die Erhebungsmethode der Beobachtung sowie experimentelle Forschungsdesigns und Verwendung von Onlineressourcen zur Erhebung der Patientenzufriedenheit eingegangen. Darüber hinaus wird den vorgestellten Erhebungsmethoden jeweils zumindest ein praktisches Beispiel zugeordnet, in welchem die jeweilige Methode angewandt wurde, um die Patientenzufriedenheit mit Hausärztinnen und -ärzten zu messen. Durch die Verbindung von Theorie und Praxis innerhalb des Kapitels soll eine umfassende Übersicht über Möglichkeiten der Patientenzufriedenheitsmessung gegeben werden, aus welcher, je nach jeweiligem Erkenntnisinteresse, eine oder mehrere passende Erhebungsmethoden ausgewählt werden können.

Bei der Auswahl der einzelnen Kapitelinhalte wurde quellen- und literaturuntersuchend vorgegangen. Anmerkungen zur diesbezüglichen Such- und Auswahlstrategie sind im Anhang ab Seite 101 zu finden.

2. Begriffliche und konzeptionelle Grundlagen

Im Folgenden werden begriffliche und konzeptionelle Grundlagen ausgeführt, welche für das Verständnis des Konstruktes der Patientenzufriedenheit, des Entstehungsprozesses von Patientenzufriedenheit und der Besonderheiten innerhalb der Arzt-Patienten-Beziehung bedeutsam sind. Dazu werden zuerst die Begriffe Zufriedenheit, Kundenzufriedenheit und Patientenzufriedenheit definiert und voneinander abgegrenzt. Danach wird zur Charakterisierung der Entstehung von Patientenzufriedenheit das Confirmation / Disconfirmation Paradigma und zur Konzeption der Faktorstruktur der Patientenzufriedenheit das Kano-Modell dargestellt. Abschließend sollen Überlegungen bzgl. des zwischenmenschlichen Verhältnisses von Ärztinnen / Ärzten und Patientinnen / Patienten dabei helfen, die besondere Bedeutung der Arzt-Patienten-Beziehung zu spezifizieren sowie Unterschiede

und Gleichartigkeiten zu anderen (Geschäfts-) Beziehungen aufzudecken.

2.1. Zufriedenheit

Zufriedenheit beschreibt einen mentalen Faktor des Wohlbefindens, welcher durch die positive Beurteilung von Gegebenheiten entsteht. Die Beurteilung wiederum entwickelt sich durch einen Vergleich der gegenwärtigen Situation mit subjektiven Vorsätzen und Erwartungen.⁸ Dementsprechend spezifiziert Zufriedenheit eine psychologische Gegebenheit, von welcher Individuen eine stärker oder schwächer ausgeprägte, aber dennoch nicht direkt vergleichbare Anschauung haben. Allen Vorstellungen gemein ist jedoch, dass der Begriff Zufriedenheit in gewissem Maße als wünschenswerter Zustand verstanden wird, welcher mit Erfüllung und Genugtuung, Genuss und Vergnügen verbunden ist.⁹ Als Folge dessen definiert Zufriedenheit das Ergebnis der Beurteilung von Gegebenheiten und Sachverhalten, Individuen und Charakteren sowie materiellen und immateriellen Symbolen, Entitäten und Gegenständen, wobei die Beurteilung positiv, negativ oder neutral ausfallen kann. Elemente mit positiver oder negativer Priorisierung sind hierbei mit einer gewissen Wertigkeit behaftet. Im Gegensatz dazu gelten neutral beurteilte Komponenten als nicht wertrelevant, da neutrale Evaluationen Indifferenz ausdrücken.¹⁰ Den stärksten Einfluss auf die subjektiv empfundene Zufriedenheit hat die spezifische Persönlichkeit, obwohl Situationen und Lebensumstände in manchen Fällen auch einen erheblichen Einfluss ausüben können. Darüber hinaus kann sich die Persönlichkeit und damit verbunden das Beurteilungsschema der subjektiven Zufriedenheit im Laufe der Zeit verändern.¹¹

Zusammenfassend kann daraus geschlossen werden, dass es sich beim Konstrukt der Zufriedenheit um einen mental-emotionalen Effekt handelt, welcher seinen Ursprung in der Bewertung von Sachverhalten hat.¹²

2.1.1. Kundenzufriedenheit

In der ökonomischen Verhaltenstheorie wird Zufriedenheit als theoretisches Gebilde wahrgenommen, um das Konsum- und Auskunftsverhalten von Verbrauchern zu konkretisieren. Diese Betrachtungsweise kann grundsätzlich drei unterschiedliche Ausprägungen annehmen. Zufriedenheit kann als Konsequenz einer mentalen Gegenüberstellung von prognostizierter und eingetretener Bedarfsbefriedigung, von prognostizierten und real erfassten Attributmerkmalen eines Produktes bzw. einer Leistung oder von optimalen und real erfassten Attributmerkmalen eines Produktes bzw. einer Leistung verstanden werden. Die Beurteilung bezieht sich auf vollständige Systeme (z.B. Organisationen) oder individuelle Leistungen von Unternehmen.¹³

⁸Vgl. Mayring (1991), S. 2007.

⁹Vgl. Scharnbacher und G. (2003), S. 5.

¹⁰Vgl. Ipsen (1978), S. 45.

¹¹Vgl. Richard und Diener (2009), S. 75.

¹²Vgl. Scharnbacher und G. (2003), S. 5.

¹³Vgl. Kirchgorg (o.J.), o.S.

Im Rahmen von nutzbringenden Tauschgeschäften mündet dementsprechend Zufriedenheit oder Unzufriedenheit in einem mental-emotionalen Effekt, welcher von der Kundin / vom Kunden ausgehend eine unternehmerische Leistung bewertet.¹⁴ Unter Kundenzufriedenheit wird folglich ein variables Gefühlsgeflecht eines oder einer Reihe von Kundinnen / Kunden verstanden, welches durch die Gegenüberstellung von vorausgesetztem Nutzen vor dem Erwerb einer Ware (bzw. Bezug einer Leistung) und dem faktischen Nutzen nach dem Erwerb entsteht. Der Nutzen einer Ware oder einer Leistung umfasst hierbei unterschiedliche Kriterien, wie bspw. Leistungsmerkmale, Rentabilität, Zeitrressourcenersparnis, Schönheit, Prestige oder Wohlbehagen.¹⁵

Zusammenfassend wird unter Kundenzufriedenheit die Konsequenz eines Standpunktes bzw. einer Gegenüberstellung von prognostizierter und erhaltener Leistung verstanden. Die Kundenforderungen werden hierbei von persönlichen Ambitionen, Angebotskenntnissen, Optionen und Auswahlmöglichkeiten im Markt sowie durch das Ansehen der Anbieterin / des Anbieters beeinflusst.¹⁶

Das Interesse an der Messung der Kundenzufriedenheit hat in der Vergangenheit kontinuierlich zugenommen. Dieser Umstand ist nicht zuletzt darauf zurückzuführen, dass mehrfach belegt werden konnte, dass erhöhte Kundenzufriedenheit den nachhaltigen Unternehmenserfolg sichert.¹⁷

2.1.2. Patientenzufriedenheit

Patientenzufriedenheit beschreibt eine Komponente des Gesundheitszustandes und gilt als elementarer Bestandteil der qualitativen Bewertung von medizinischen Leistungen.¹⁸ Obwohl dem Begriff Patientenzufriedenheit keine einheitliche Definition zugrunde liegt,¹⁹ wird zumindest das Konzept der Patientenzufriedenheit durch einen Vergleichsprozess von erwarteter und erhaltener Leistung definiert.²⁰ Patientenzufriedenheit kann demnach als persönliche Wahrnehmung einer Gesundheitseinrichtung durch die Patientin / den Patienten betrachtet werden. Hierbei wird nicht nur das Niveau der Schlüsselperformanz, der ärztlichen Behandlung, sondern auch die Gesamterfahrung des Einrichtungsbesuches in die Bewertung miteinbezogen.²¹

Patientenzufriedenheit kann als Bewertungsprozess der Erwartungserfüllung verstanden werden, wobei die Erwartungen an den Gesundheitsbereich viel stärker an Wünsche und Vorstellungen gebunden sind als bspw. die Erwartungen an den Einzelhandelsbereich. Diese Gegebenheit findet ihre Begründung hauptsächlich darin, dass der Gesundheitsbereich mental komplexer wahrgenommen wird als andere Bereiche und daher die Einzelkomponentenerwartung nicht

so stark ausgeprägt ist.²² Während die positive Patientenzufriedenheitsbewertung also individuell gebildet wird und dadurch beinahe unbegrenzte Faktoren und Komponenten beinhalten kann, entsteht negative Patientenzufriedenheit im Allgemeinen nur unter zwei voneinander abweichenden Voraussetzungen. Einerseits durch eine Begebenheit, die nicht entstehen hätte sollen, andererseits durch einen prognostizierten Zustand, welcher nicht eingetreten ist.²³

Obwohl sich die Rolle der Patientin / des Patienten als Individuum in der Vergangenheit stark verändert hat und diese / dieser im Gesundheitswesen immer stärker als Kundin / Kunde wahrgenommen wird,²⁴ kann nicht davon ausgegangen werden, dass es sich bei Patientenzufriedenheit um das gänzlich gleiche Konstrukt handelt wie bei Kundenzufriedenheit.²⁵ Patientinnen und Patienten können nicht als bloße Konsumentinnen und Konsumenten von Gesundheitsleistungen betrachtet werden, da diese zumeist mittelbar oder unmittelbar an der gesundheitlichen Wertschöpfung beteiligt sind. Das bedeutet, dass Patientinnen und Patienten einen deutlich stärkeren Einfluss auf das positive oder negative Behandlungsergebnis haben als Kundinnen und Kunden anderer Sektoren. Als Folge dessen ist davon auszugehen, dass zumindest ein Teil der wahrgenommenen Leistung direkt oder indirekt durch die Patientin / den Patienten selbst beeinflusst wird.²⁶ Ferner sind bspw. erheblich erkrankte Patientinnen und Patienten in ihrer autonomen Kundenstellung eingeschränkt und oft auch nicht mehr fähig, ihre individuelle Zufriedenheit zu beurteilen.²⁷ Des Weiteren ist der Beginn und der Ausgang einer ärztlichen Behandlung für Patientinnen und Patienten nicht immer frei wählbar und im Notfall haben sie weder Einfluss auf den Behandlungsort noch auf die Behandlungssituation.²⁸ Da sich die Kundenzufriedenheit in der Beurteilung einer Ware oder Dienstleistung ausdrückt²⁹ und dieses Beurteilungsschema in gewisser Weise auch auf die Patientenzufriedenheit übertragbar erscheint,³⁰ ist jedoch davon auszugehen, dass es sich bei Patientenzufriedenheit um ein abgewandeltes Konzept der Kundenzufriedenheit handelt.³¹ Das bedeutet, dass Patientenerwartungen, welche in ausreichendem Ausmaß erfüllt werden, die subjektiv empfundene Zufriedenheit erhöhen, was mittelfristig nachweislich auch zu elaborierten Behandlungsergebnissen führt. Dementsprechend resultieren Patientenzufriedenheit und Patientenbindung in dem von der Patientin / vom Patienten empfundenen Praxis- und Behandlungsniveau.³²

Patientenzufriedenheit ist die Konsequenz einer Gegenüberstellung von erwarteter und erhaltener Leistung.³³ Die

¹⁴Vgl. Scharnbacher und G. (2003), S. 5.

¹⁵Vgl. Schneider et al. (2008), S. 177.

¹⁶Vgl. Karner (2004), S. 619.

¹⁷Vgl. Homburg und Bucerius (2006), S. 55.

¹⁸Vgl. Donabedian (1988), S. 1746.

¹⁹Vgl. Al-Abri und Al-Balushi (2014), S. 3.

²⁰Vgl. Sitzia und Wood (1997), S. 1840.

²¹Vgl. Thill (2008), S. 3.

²²Vgl. Williams (1994), S. 509 f.

²³Vgl. Wüthrich-Schneider (2000a), S. 1046.

²⁴Vgl. Neugebauer und Porst (2001), S. 1.

²⁵Vgl. Wengle (2000), S. 2440.

²⁶Vgl. Frodl (2011), S. 124.

²⁷Vgl. Rogler (2009), S. 1009.

²⁸Vgl. Kick (2006), S. A1207.

²⁹Vgl. Stauss und Seidel (2014), S. 59.

³⁰Vgl. Neugebauer und Porst (2001), S. 1.

³¹Vgl. Wengle (2000), S. 2440.

³²Vgl. Theis (2007), S. 8.

³³Vgl. Williams (1994), S. 509 f.

Erwartung selbst bildet sich hierbei subjektiv individuell und wird durch eine Vielzahl von Faktoren wie bspw. vorherigen Erfahrungen, Umwelt- und Umgebungseinflüssen oder Wünschen beeinflusst.³⁴ Auch soziodemografische Faktoren, persönliche Informationsmöglichkeiten (z.B. digitale oder verbale Empfehlungen) oder individuelle Bedürfnisse können die gebildete Erwartungshaltung nachhaltig beeinflussen.³⁵ Vor diesem Hintergrund kann einerseits festgestellt werden, dass die individuelle Erwartungshaltung von Patientinnen und Patienten dazu führt, dass Einflüsse, welche bei einer Person Zufriedenheit auslösen, bei einer anderen Person zu Unzufriedenheit führen können. Andererseits kann argumentiert werden, dass Erwartungen sich je nach Situation im Zeitverlauf verändern.³⁶ Insbesondere im medizinischen Bereich ist die Erwartungshaltung von Patientinnen und Patienten elementar, da diese, erheblichen Einfluss auf den Genesungserfolg ausüben kann.³⁷ Aus diesem Grund ist es unumgänglich, die Patientenerwartungen zu erkennen und in den Behandlungsprozess miteinzubeziehen, um nachhaltige Behandlungserfolge zu erzielen.³⁸

Ein Problemfeld in diesem Zusammenhang kann durch unrealistische Patientenerwartungen, welche durch fehlende oder fehlerhafte Informationen, Perspektiven oder Interpretationen hervorgerufen werden, entstehen. Diese mindern in der Regel den Behandlungserfolg und führen häufig zu Unzufriedenheit.³⁹ Je höher hingegen der Grad an Informationen, Perspektiven oder Interpretationen ist, desto eher nähert sich die Patientenerwartung dem realistisch durchführbaren Behandlungsprozess. Begünstigt wird die realistische Patientenerwartung infolgedessen durch umfassende Informationsmöglichkeiten, verständliche Diagnosen, realistische Prognosen und universelle Aufklärung über die Wirkung von Medikamenten.⁴⁰

Ein weiteres einflussnehmendes Kriterium auf die Patientenzufriedenheit ist die persönliche Präferenz. In der ökonomischen Verhaltenstheorie wird unter Präferenz die persönliche Beurteilung eines Güterbündels A gegenüber eines Güterbündels B hinsichtlich des zu erwartenden Grades an Bedürfnisbefriedigung verstanden. Die Präferenzen eines ökonomischen Elementes beschreiben demnach das Resultat eines Beurteilungs- und Begutachtungsvorganges, welches unter Miteinbeziehung von suffizientem, subjektivem Wissen gebildet wird und innerhalb einer gewissen Zeitspanne bestehen bleibt.⁴¹

Auch die Präferenzen von Patientinnen und Patienten bilden sich individuell und variieren im Zeitverlauf. Einfluss auf die Entstehung von Präferenzen haben wiederum soziodemografische Faktoren, wobei Kriterien, wie bspw. die soziale Schicht, die Entwicklung von Präferenzen stark beein-

flussen. Da niedergelassene Ärztinnen und Ärzte im Gegensatz zu jenen in öffentlichen oder privaten Kliniken häufiger in Kontakt mit denselben Patientinnen und Patienten treten, wird insbesondere von dieser Berufsgruppe erwartet, dass die jeweiligen Patientenpräferenzen erkannt, verstanden und innerhalb des Behandlungsprozesses berücksichtigt werden.⁴²

Die Kenntnis der Präferenzen der Patientinnen und Patienten verringert einerseits die mit der Behandlung verbundenen Risiken durch umfassendere operable Behandlungsmöglichkeiten, hat sich aber andererseits auch zur Forderung der Patientinnen und Patienten gegenüber Ärztinnen und Ärzten entwickelt.⁴³ Der stärkere Fokus auf die Patientenpräferenzen zeigt sich auch in der Verschiebung von der paternalistischen Medizin zur Patientenpartnerschaft, bei welcher sich Ärztin / Arzt und Patientin / Patient auf Augenhöhe begegnen. Obwohl es grundsätzliche Unterstützung für den Miteinbezug von Patientenpräferenzen gibt, sind noch viele Umsetzungsschritte erforderlich, um die Wünsche und Vorlieben von Patientinnen und Patienten generell in das Zentrum des Behandlungsprozesses zu rücken und damit umfassende patientenorientierte Pflege sicherzustellen.⁴⁴

2.2. Entwicklung von Kunden- und Patientenzufriedenheit

Der Entwicklungsprozess der Patientenzufriedenheit unterliegt einer Vielzahl an Charakterisierungsversuchen. Ein in dieser Hinsicht häufig angewandtes Modell ist das Confirmation / Disconfirmation Paradigma (C/D-Paradigma).⁴⁵ Das C/D-Paradigma definiert das Basiskonzept zur Darstellung des Entstehungsprozesses von Kundenzufriedenheit und ist durch eine Vielzahl psychologischer Schemata, welche ausführliche theoretische Ansätze zur exakten Definition individueller Faktoren des Konzeptes bieten, fundiert. Als Beispiel kann hier die Assimilations-Kontrast-Theorie herangezogen werden, welche beschreibt, wie der Niveauunterschied von prognostizierter und wahrgenommener Leistung im Nachhinein angepasst wird.⁴⁶

Um die Struktur der einflussnehmenden Patientenzufriedenheitsfaktoren näher zu bestimmen, hat sich hingegen das Kano-Modell durchgesetzt,⁴⁷ welches auch bei der Bestimmung der Faktorstruktur der Kundenzufriedenheit häufig Anwendung findet. Das Kano-Modell beruht, wie auch schon das C/D-Paradigma, auf theoretischen psychologischen Ansätzen, welche größtenteils als anerkannt gelten. Zu nennen wäre hier die Mehrfaktorentheorie der Kundenzufriedenheit, welche versucht, Zufriedenheit als Ergebnis der Bewertung unterschiedlicher Faktoren darzustellen.⁴⁸

2.2.1. Confirmation / Disconfirmation Paradigma

Wie Abbildung 1 grafisch darstellt, entwickelt sich nach diesem Modell Zufriedenheit unter der Voraussetzung, dass

³⁴Vgl. Crow et al. (2002), S. 8.

³⁵Vgl. Frodl (2011), S. 119 f.

³⁶Vgl. Crow et al. (2002), S. 11.

³⁷Vgl. Linde et al. (2007), S. 270.

³⁸Vgl. Tahir et al. (2012), S. 2025.

³⁹Vgl. K.P. (2004), S. 191.

⁴⁰Vgl. Bhana et al. (2015), S. 130 f.

⁴¹Vgl. Piekenbrock (o.J.), o.S.

⁴²Vgl. Lin et al. (2012), S. 55.

⁴³Vgl. Bowling und Ebrahim (2001), S. i2.

⁴⁴Vgl. Laidsaar-Powell et al. (2014), S. 104.

⁴⁵Vgl. Thompson und Yarnold (1995), S. 1061.

⁴⁶Vgl. Nerdinger et al. (2015), S. 131.

⁴⁷Vgl. C. und C. (2013), S. 246.

⁴⁸Vgl. Hölzing (2008), S. 77 ff.

eine Kundin / ein Kunde ihre / seine prognostizierten Erwartungen (Vergleichsstandard) ihren / seinen gegenwärtigen Produkterfahrungen (wahrgenommenes Leistungsniveau) gegenüberstellt.⁴⁹ Erfüllt das wahrgenommene Leistungsniveau den Vergleichsstandard, so werden die Erwartungen bestätigt. Dieser Vorgang führt dem Konzept entsprechend zu Konfirmation (Zufriedenheit).⁵⁰ Wird der Vergleichsstandard vom wahrgenommenen Leistungsniveau übertroffen, so werden auch die Erwartungen übertroffen, was zu positiver Diskonfirmation (hoher Zufriedenheit) führt.⁵¹ Dem gegenüber steht die negative Diskonfirmation (Unzufriedenheit), welche aus einem unterhalb des Vergleichsstandards wahrgenommenen Leistungsniveaus resultiert.⁵²

Obwohl sich das C/D-Paradigma im deutschen Sprachraum als Erklärungsmodell von Zufriedenheit durchgesetzt hat, muss unter bestimmten Voraussetzungen daran gezweifelt werden, dass es allgemeine Gültigkeit besitzt. Einerseits kann nicht davon ausgegangen werden, dass die Erfüllung von negativ empfundenen Prognosen Zufriedenheit erzeugt. Es ist also anzunehmen, dass Kundinnen und Kunden, welche eine widrige Performanz erwarten und diese auch wahrnehmen, als Konsequenz der Bestätigung ihrer Erwartung nicht intuitiv zufrieden sind. Demnach würde das C/D-Paradigma nur für Kundinnen und Kunden gelten, welche im Vorfeld positive Prognosen anstellen. Andererseits kann sich der Abstand zwischen prognostizierter und wahrgenommener Leistung durch Erlebnisse und Kenntnisse der Kundinnen / Kunden im Zeitverlauf verringern. Im Fall von positiver Diskonfirmation würde dies dem Modell entsprechend bedeuten, dass sich die hohe Zufriedenheit der Zufriedenheit annähert – das Zufriedenheitslevel also nachhaltig sinkt. Dieser Erfahrungskurveneffekt führt jedoch nicht zwingend zu reduzierter Zufriedenheit.⁵³

Trotz der Kritikpunkte wird das C/D-Paradigma mehrheitlich als konzeptionelles Modell angewandt, um die Bestandteile eines Zufriedenheitsbewertungsprozesses grafisch darzustellen.⁵⁴ Unter anderem ist dies darauf zurückzuführen, dass sich das Modell zu einem nützlichen Hilfsinstrument zur Beeinflussung der Kundenzufriedenheit entwickelt hat. In diesem Zusammenhang lassen sich drei Ansatzpunkte ablesen. Einerseits kann die unternehmerische Leistungsqualität gesteigert werden, des Weiteren kann durch entsprechende Kommunikationsmaßnahmen das Qualitätsbewusstsein der Kundinnen und Kunden nachhaltig positiv beeinflusst werden und durch die Modellanwendung besteht die Möglichkeit, die Erwartungen der Kundinnen und Kunden selbst zu beeinflussen.⁵⁵

2.2.2. Das Kano-Modell

Das Kano-Modell bietet eine relativ einfache Möglichkeit, die Faktorstruktur der Entwicklung von Zufriedenheit grafisch darzustellen. Hierbei wird einerseits der Grad an Zufriedenheit bei der Betrachtung miteinbezogen, andererseits werden auch die einzelnen Faktoren innerhalb der Attribute Grundanforderung, Leistungsanforderung und Begeisterungseigenschaft gruppiert.⁵⁶

Grundanforderungen beinhalten jene Leistungseigenschaften, welche zu Unzufriedenheit führen, sollte die Prognose nicht mit der wahrgenommenen Leistung übereinstimmen. Wird die prognostizierte Leistung erreicht oder positiv übertroffen, führt dieser Umstand nicht zu Zufriedenheit, löst aber keine Unzufriedenheit aus. Unter Grundanforderungen sind demnach jene Voraussetzungen zu verstehen, welche für die Entwicklung von Zufriedenheit zwar erforderlich, jedoch alleine nicht ausreichend sind.⁵⁷

Im Zentrum des Bedürfniserfüllungsprozesses stehen die Leistungsanforderungen. Diese beschreiben die Kernleistung und damit einhergehende Begleitfaktoren. Mit steigendem Erfüllungsgrad steigt auch die Zufriedenheit. Im Umkehrschluss sinkt die Zufriedenheit bei unzureichender Erfüllung.⁵⁸

Im Gegensatz dazu mündet die Erfüllung bzw. das Übertreffen der erwarteten Leistung im Bereich der Begeisterungseigenschaften stets in hoher Zufriedenheit, während ein niedriges Leistungslevel innerhalb dieser Kategorie nicht zwingend zu Unzufriedenheit führt. Dieses Szenario kommt zustande, da Kundinnen und Kunden in Bezug auf Begeisterungseigenschaften in der Regel keine besonders hohen Erwartungen haben.⁵⁹ In Abbildung 2 sind die Komponenten des Kano-Modells zusammengefasst dargestellt.

Die Kategorisierung von Faktoren als Grund- oder Leistungsanforderung bzw. als Begeisterungsfaktoren steht im Zusammenhang mit der kundenspezifischen Einstellung und unterscheidet sich auch erheblich in Hinblick auf die Betrachtung unterschiedlicher Zielgruppen. Da sich die Kundeneinstellung im Zeitverlauf ändert, kann auch die diesbezügliche Kategorisierung von der Änderung betroffen sein. Faktoren, welche aktuell Begeisterung auslösen, können mittelfristig zu einem Bestandteil der Kernleistung avancieren und langfristig als Grundvoraussetzung gelten.⁶⁰

Obwohl das Kano-Modell der Kundenzufriedenheit weitgehend Zuspruch findet, bestehen Kritikpunkte hinsichtlich der deduktiven Schlussfolgerung. Abgesehen davon wird das Modell aber sowohl in wissenschaftlicher als auch in praktischer Hinsicht als Instrument zur Identifikation der Zufriedenheitsfaktorstruktur anerkannt.⁶¹

⁴⁹Vgl. Homburg und Krohmer (2006), S. 44.

⁵⁰Vgl. Töpfer (2008a), S. 50.

⁵¹Vgl. Homburg und Krohmer (2006), S. 44.

⁵²Vgl. Homburg und Rudolph (1997), S. 38 f.

⁵³Vgl. Bartikowski (2002), S. 11.

⁵⁴Vgl. Bösemer (2015), S. 13.

⁵⁵Vgl. Homburg und Krohmer (2006), S. 47.

⁵⁶Vgl. Oliver (2010), S. 158.

⁵⁷Vgl. Matzler et al. (2004), S. 19.

⁵⁸Vgl. Töpfer (2008c), S. 197.

⁵⁹Vgl. Homburg und Krohmer (2006), S. 512.

⁶⁰Vgl. Matzler und Bailom (2004), S. 287.

⁶¹Vgl. Hölzing (2008), S. 203.

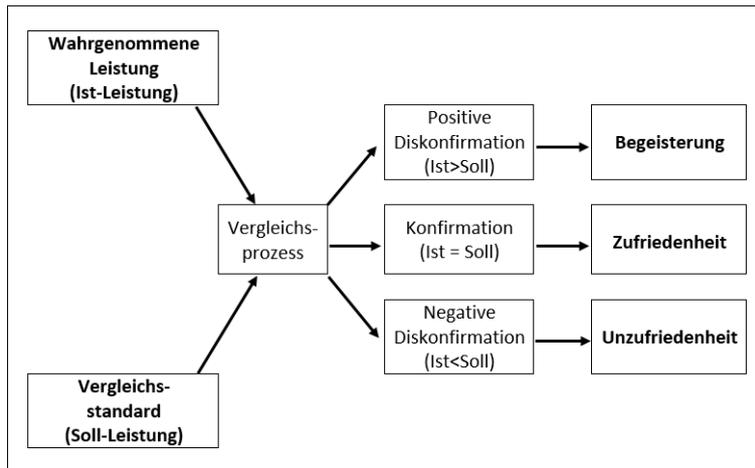


Abbildung 1: Modellkomponenten des C/D-Paradigmas; Quelle: In Anlehnung an , S. 53.

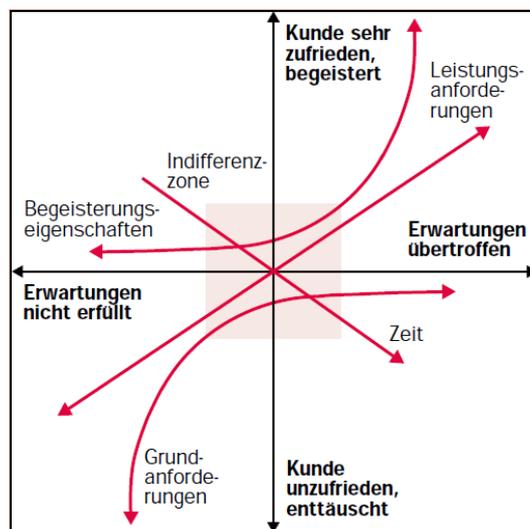


Abbildung 2: Das Kano-Modell der Kundenzufriedenheit; Quelle: Bailom et al. (1998), S. 48.

2.3. Besonderheiten in der Arzt-Patienten-Beziehung

Die Arzt-Patienten-Beziehung ist eine der bedeutsamsten Kerngrößen in der Betrachtung gesundheitsspezifischer Leistungen und daher von hohem sozialwissenschaftlichen, ethnischen und ökonomischen Interesse.⁶² Die Leistungen, welche innerhalb der Arzt- Patienten-Beziehung erbracht werden, können hierbei in bestimmtem Umfang durchaus als Dienstleistungen verstanden werden und sind in der Regel auch durch deren Eigenschaften gekennzeichnet. Da die interne und externe körperliche und geistige Unversehrtheit für Patientinnen und Patienten einen sehr hohen Stellenwert einnimmt, in vielen Fällen sogar als lebensnotwendig angesehen wird, kann jedoch nicht davon ausgegangen werden, dass Dienstleistungen im Gesundheitsbereich vollkommen mit materiellen Dienstleistungen vergleichbar sind. Darüber hinaus bewegen sich Gesundheitsdienstleistungen innerhalb eines subjektiven Bezugsrahmens, wodurch die

Arzt-Patienten-Beziehung auch aus Patientinnen- und Patientensicht an Bedeutung gewinnt.⁶³

Ein maßgebliches Element in dieser Hinsicht ist die Ungleichmäßigkeit der Arzt-Patienten- Beziehung, welche besonders hinsichtlich der Informationsdiskrepanz deutlich wird. Das bedeutet, dass das Verhältnis zwischen Ärztinnen / Ärzten und Patientinnen / Patienten durch relativ hohe Wissens- und Kenntnisunterschiede gekennzeichnet ist. Diese Sichtweise hat grundsätzlich zwei Ausprägungen. Einerseits verfügt die behandelnde Ärztin / der behandelnde Arzt im Regelfall über höhere Bereichs- und Methodenkompetenz im medizinischen Bereich, was dazu führt, dass Patientinnen und Patienten in gewisser Weise auf die medizinischen Fachkenntnisse der behandelnden Ärztin / des behandelnden Arztes vertrauen müssen. Andererseits haben Patientinnen und Patienten einen Wissensvorsprung hinsichtlich des persönlichen Krankheitsbildes und der Symptome und wis-

⁶²Vgl. Bestmann und Verheyen (2010), S. 7.

⁶³Vgl. Mehmet (2011), S. 5.

sen häufig bereits aus Erfahrung, welche Behandlungs- und Therapiemethoden am ehesten erfolgsversprechend wirken. Die Arzt-Patienten-Beziehung ist demnach dadurch gekennzeichnet, dass auf beiden Seiten Informations- und Kompetenzunterschiede bestehen, wobei die Ärztinnen und Ärzte den medizinischen Bereich besser beurteilen können und Patientinnen und Patienten Spezialisten im Bereich ihrer subjektiven Existenz sind.⁶⁴

Diese Erkenntnis hat dazu geführt, dass Patientinnen und Patienten heutzutage keine reinen LeistungsempfängerInnen mehr sind, sondern häufig aktiv in den ärztlichen Entscheidungsprozess miteinbezogen werden, wobei der Beteiligungsgrad naturgemäß variiert. Dadurch wird auch das klassisch paternalistische Ärztebild, bei welchem Ärztinnen und Ärzte eine elterliche Rolle einnehmen, durch Behandlungspartnerschaft und ein wachsendes Maß an Patientenautonomie substituiert.⁶⁵ Durch diese Entwicklung gewinnt die Patientin / der Patient als Individuum an Bedeutung und traditionelle Rollenbilder werden durch neue Ansätze ersetzt. Aus diesem Grund werden Patientinnen und Patienten gegenwärtig häufig als LeistungsnutzerInnen oder Leistungskonsumentinnen und -konsumenten betrachtet.⁶⁶ Moderne Betrachtungsweisen fokussieren sich oftmals generell auf die ökonomische Sichtweise der Arzt-Patienten-Beziehung, bei welcher gewöhnlich die Arztpraxis als dienstleistendes Unternehmen in einer polypolistischen Marktwirtschaft dargestellt wird und die Patientin / der Patient als Kundin / Kunde in diesem vollkommenen Markt fungiert.⁶⁷ Untersuchungen in diesem Bereich haben allerdings gezeigt, dass die Rollendefinition von Patientinnen und Patienten durchaus kontextabhängig schwanken kann. So werden EmpfängerInnen von Gesundheitsleistungen je nach Situation und Behandlungsumfeld als Patientinnen / Patienten, NutzerInnen, Konsumentinnen / Konsumenten oder Kundinnen / Kunden betrachtet. Abgesehen von Extrem- bzw. Notfallsituationen hat die Rolle der Patientin / des Patienten innerhalb der Arzt-Patienten-Beziehung jedoch situationsunabhängig an Bedeutung gewonnen.⁶⁸

3. Einflussfaktoren auf die Patientenzufriedenheit

Patientenzufriedenheit mit der zur Verfügung stehenden ärztlichen Versorgung ist in vielerlei Hinsicht von großer Bedeutung. Aus diesem Grund ist es elementar, auftretende Problemfelder und Faktoren zeitgerecht zu identifizieren, um negativen Entwicklungen rechtzeitig entgegenwirken zu können.⁶⁹ Da Patientenzufriedenheit jedoch von einer Vielzahl mehr oder weniger komplexer Faktoren abhängig ist, kann sich die Identifikation jener Problembereiche unter Umstän-

den sehr anspruchsvoll gestalten.⁷⁰ Aus diesem Grund werden im Folgenden verschiedene Modelle zur Grobkategorisierung von Patientenzufriedenheit mit niedergelassenen Ärztinnen und Ärzten dargestellt und unterschiedliche, beeinflussbare Patientenzufriedenheitsfaktoren erörtert.

3.1. Modelle zur Kategorisierung von Patientenzufriedenheit

Fan et al. unterscheiden in ihrer Studie aus dem Jahr 2005 patientenspezifische, anbieterspezifische und organisationale Einflussfaktoren auf die Patientenzufriedenheit.⁷¹ Diese Kategorisierung folgt auch dem Modell von Jackson et al., in welchem davon ausgegangen wird, dass Patientenzufriedenheit die Summe aus dem bewerteten Behandlungsergebnis, der wahrgenommenen Servicequalität, demografischen Aspekten und dem gesamten Behandlungsprozess darstellt.⁷² Unter patientenspezifischen Einflussfaktoren wird in diesem Zusammenhang die persönliche Patientenwahrnehmung,⁷³ aber auch sozioökonomische und demografische Aspekte wie das Geschlecht, das Alter, der Bildungsgrad, der Gesundheitsstatus oder die soziale Stellung verstanden.⁷⁴ Währenddessen umfassen anbieterspezifische Faktoren jene Elemente, welche die Hausärztin / den Hausarzt selbst betreffen und dementsprechend als direkt beeinflussbar gelten. Dazu zählen bspw. fachliche oder charakteristische Arzteigenschaften, das Praxispersonal, die Praxisausstattung, Komfort und Umgebungscharakteristika. Organisationale Faktoren definieren Fan et al. als Menge der Arztbesuche innerhalb von 12 Monaten und der geografischen Entfernung zur Arztpraxis und legen sich damit wiederum auf durch die Ärztin / den Arzt nicht oder nur schwer beeinflussbare Faktoren fest.⁷⁵ Die einzelnen Modellkomponenten sind in Abbildung 3 übersichtlich zusammengefasst.

Ein weiteres Modell zur Grobkategorisierung von Patientenzufriedenheit wurde im Zuge der QUALICOPC (Quality and Costs of Primary Care in Europe) Studie erstellt. Die QUALICOPC Studie vergleicht Leistungen von europäischen Primärpflegesystemen in Bezug auf Qualität, Eigenkapital und Kostenstruktur.⁷⁶ Um eine umfassende Datenmenge zu gewährleisten, wurden insgesamt 31 Länder in Europa sowie Kanada, Israel und Australien untersucht, wobei die Untersuchungsgebiete außerhalb Europas selbstständig an der Untersuchung teilnahmen.⁷⁷ Insgesamt nahmen demnach 34 Länder an der QUALICOPC Studie teil, bei welcher als Querschnittsuntersuchung Daten von Hausärztinnen und -ärzten und deren Patientinnen und Patienten gesammelt wurden.⁷⁸ Schäfer et al. definieren innerhalb der QUALICOPC Studie die drei Hauptkategorien Systemlevel, Dienstleistungslevel und Patientenwahrnehmungslevel, wobei die

⁷⁰Vgl. Alberto Sánchez et al. (2014), S. 153.

⁷¹Vgl. Fan et al. (2005), S. 227 f.

⁷²Vgl. Jackson et al. (2001), S. 610.

⁷³Vgl. Fan et al. (2005), S. 227.

⁷⁴Vgl. Hekkert et al. (2009), S. 74.

⁷⁵Vgl. Fan et al. (2005), S. 227.

⁷⁶Vgl. Schäfer et al. (2011), S. 1.

⁷⁷Vgl. Cohidon et al. (2015), S. 2.

⁷⁸Vgl. Wong et al. (2015), S. 2.

⁶⁴Vgl. Bestmann und Verheyen (2010), S. 7.

⁶⁵Vgl. Kaba und Sooriakumaran (2007), S. 64 f.

⁶⁶Vgl. Dierks und W. (2001), S. 796.

⁶⁷Vgl. Bestmann und Verheyen (2010), S. 7.

⁶⁸Vgl. Wildner (2011), S. 398.

⁶⁹Vgl. Thayaparan und Mahdi (2013), S. 1.

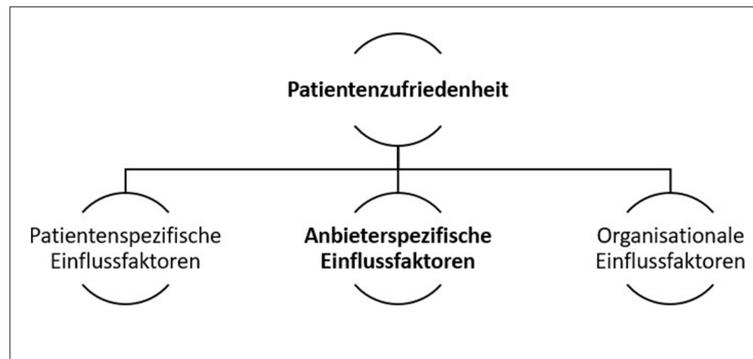


Abbildung 3: Einflussfaktoren auf die Patientenzufriedenheit nach Fen et al.; Quelle: In Anlehnung an Fan et al. (2005), S. 227 f.

Kategorien sowohl Attribute beinhalten, welche die Patientenzufriedenheit beeinflussen, als auch solche, welche lediglich das jeweilige Gesundheitssystem darstellen. Den Kategorien wurden ferner einzelne Bestandteile der Erstversorgung zugeordnet. So definiert das Systemlevel das Design und die Organisation von Erstversorgungssystemen und besteht aus Attributen wie der Finanzierung, gesetzlicher Rahmenbedingungen und Ressourcen. Das Dienstleistungslevel hingegen summiert die hausärztlichen Qualitätskriterien wie die Qualität und den Umfang der Pflege, die Kontinuität, Vorsorgeaktivitäten oder die Orientierung auf das Gemeinwohl; das Patientenwahrnehmungslevel beschreibt die durch die Patientin / den Patienten wahrgenommene Qualität und beinhaltet Eigenschaften wie den Zugang zu Primärversorgung oder den wahrgenommenen Komfort. Obwohl die vordefinierten Kategorien, wie auch die Untersuchung selbst, nicht direkt die Messung von Patientenzufriedenheit mit Hausärztinnen und -ärzten anstreben, entwickeln sich aus ihnen innerhalb der Untersuchung schließlich Systemziele, welche im Sinne von Qualität und Patientenzufriedenheit erreicht werden sollen. Diese sind Zugang zur Primärversorgung, Kosten und Effizienz, Serviceprozessqualität und wahrgenommene Serviceprozessqualität.⁷⁹ Die Zufriedenheitsfaktorenkategorisierung von Schäfer et al. wird angewandt, um Einflussfaktoren auf die Patientenzufriedenheit, insbesondere in europäischen Ländermärkten, zu untersuchen.⁸⁰ Ein weiterer Anknüpfungspunkt ist der Vergleich der Prozessqualitätsindikatoren von Hausärztinnen und -ärzten in unterschiedlichen Ländermärkten wobei hier Ähnlichkeiten und Unterschiede innerhalb und zwischen den einzelnen Ländern und die Gründe für die Unterschiede, bspw. Eigenschaften der Ärztin / des Arztes, untersucht werden.⁸¹ Abbildung 4 fasst die Patientenzufriedenheitskategorien nach Schäfer et al. zusammen.

Das dritte Modell zur Grobkategorisierung von Patientenzufriedenheit, welches hier vorgestellt wird, wurde von Kontopantelis et al. aufbauend auf die Auswertung der Er-

gebnisse der General Practice Access Survey 2007 entwickelt. Innerhalb der Studie wird zwischen Patientenebene, Praxisebene und Umgebungsebene unterschieden. Die Patientenebene beschreibt hierbei etliche demografische Patienten Aspekte, wie das Geschlecht oder die Altersgruppe, aber auch Faktoren, wie das momentane Anstellungs- bzw. Arbeitsverhältnis, die Wochenarbeitszeit sowie die ländliche oder urbane Wohnumgebung. Die Praxisebene spezifiziert Patientenzufriedenheitsdimensionen, wie die Praxisgröße und Öffnungszeiten und die Umgebungsebene summiert regionale Aspekte, wie das prozentuale Gesundheitspersonal pro 100.000 Einwohner oder andere regionale Einflussfaktoren.⁸² Die Studie von Kontopantelis et al. wird häufig herangezogen, um Effekte der Betreuungskontinuität zu verdeutlichen.⁸³ Weitere beliebte Ansatzpunkte, welche sich durch die Untersuchung von Kontopantelis et al. gut darstellen lassen, sind der Zusammenhang zwischen Patientenzufriedenheit und Erreichbarkeit oder Wartezeit auf einen Termin und die unterschiedlichen Patientenzufriedenheitslevel in Abhängigkeit von Praxispersonal und Praxisgröße.⁸⁴ Die diesbezüglichen Modellkomponenten sind in Abbildung 5 zusammengefasst.

Die Darstellung der Kategorisierung von Fen et al., Schäfer et al. und Kontopantelis et al. soll verdeutlichen, welchen Umfang die Bestimmung der Faktorstruktur von Patientenzufriedenheit mit Hausärztinnen und -ärzten annehmen kann. Allen Untersuchungen gemein ist, dass die Überkategorien entweder durch die Ärztin / den Arzt direkt beeinflussbare Faktoren oder ausschließlich indirekt oder nicht beeinflussbare Faktoren beinhalten. In weiterer Folge wird ausschließlich auf beeinflussbare Patientenzufriedenheitsfaktoren eingegangen, da diese in der Regel direkt gesteuert werden können und somit für Hausärztinnen und -ärzte das größte Erkenntnisinteresse im Bereich der Patientenzufriedenheit beinhalten. Beeinflussbare Faktoren sind bei Fan et al. in der Kategorie anbieterspezifische Einflussfaktoren zu finden, Schäfer et al. summieren diese innerhalb des Pati-

⁷⁹Vgl. Schäfer et al. (2011), S. 4.

⁸⁰Vgl. Eide et al. (2016), S. 7.

⁸¹Vgl. Pavlič et al. (2015), S. 1.

⁸²Vgl. Kontopantelis et al. (2010), S. 3 f.

⁸³Vgl. Aboulghate et al. (2012), S. 574.

⁸⁴Vgl. Casalino et al. (2014), S. 7.

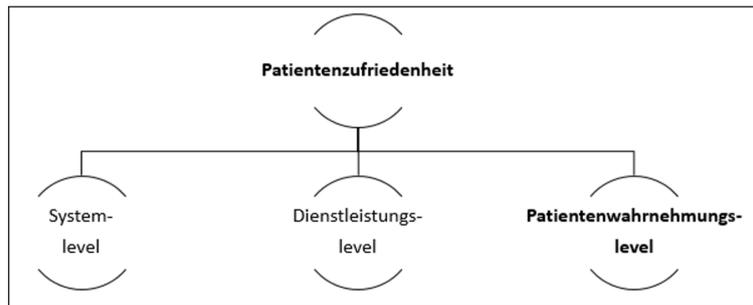


Abbildung 4: Einflussfaktoren auf die Patientenzufriedenheit nach Schäfer et al.; Quelle: In Anlehnung an Schäfer et al. (2011), S. 4 f.

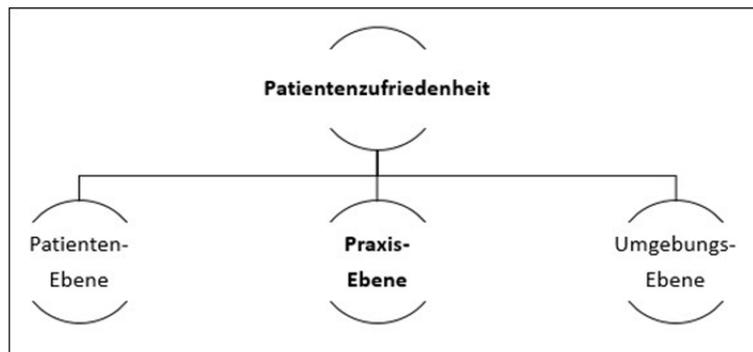


Abbildung 5: Einflussfaktoren auf die Patientenzufriedenheit nach Kontopantelis et al.; Quelle: In Anlehnung an Kontopantelis et al. (2010), S. 3 f.

entenwahrnehmungslevels und Kontopantelis et al. ordnen direkt beeinflussbare Faktoren im Zusammenhang mit Patientenzufriedenheit der Praxisebene unter. Wie die drei Modelle zeigen, wird also häufig davon ausgegangen, dass sich Patientenzufriedenheit aus unterschiedlichen, beeinflussbaren oder nicht beeinflussbaren Kategorien zusammensetzt. Unterschiedliche Ansätze gibt es jedoch bei der Faktorzuordnung, d.h. bei der Frage danach, aus welchen Faktoren sich die jeweiligen Kategorien zusammensetzen.

3.2. Ausgewählte Patientenzufriedenheitsfaktoren

Shelton unterscheidet im Bereich der beeinflussbaren Patientenzufriedenheitsfaktoren die Kategorien Zugang und Erreichbarkeit, Komfort, Kommunikation, wahrgenommene Behandlungsqualität, Fürsorge sowie Einrichtung und technische Ausstattung.⁸⁵ Andere Autoren gehen noch eine Stufe tiefer und untersuchen einzelne beeinflussbare Faktoren oder Faktorenbündel. So haben Kim et al. in ihrer Studie aus dem Jahr 2004 versucht, den Zusammenhang von Empathie gegenüber der behandelnden Ärztin / dem behandelnden Arzt und Patientenzufriedenheit darzustellen und dabei ein Modell entwickelt, in welchem sich auch der Zusammenhang mit dem Patientenbeschwerdeverhalten abbilden lässt. Die Ergebnisse dieser Untersuchung deuten darauf hin, dass die Patientenwahrnehmung in Bezug auf Empathie gegenüber

Ärztinnen und Ärzten die Patientenzufriedenheit direkt signifikant beeinflusst.⁸⁶ Zu einem ähnlichen Ergebnis in dieser Untersuchungskategorie gelangen auch Derksen et al., die in ihrer Literaturübersicht insgesamt 964 Patientenzufriedenheitsstudien auf den Zusammenhang von Empathie und Patientenzufriedenheit hin untersucht haben. Obwohl, bedingt durch die strengen Ausschlusskriterien, nur ein kleiner Prozentsatz der Studien in die Hauptuntersuchung miteinbezogen wurde, konnte bei den verbliebenen stets ein starker Zusammenhang zwischen Empathie gegenüber der behandelnden Hausärztin / dem behandelnden Hausarzt und subjektiv empfundener Patientenzufriedenheit nachgewiesen werden.⁸⁷ Andere Autoren fokussieren sich auf spezifischere Faktoren wie die Kontinuität, also den Einfluss von häufigen Hausarztwechseln. So konnten bspw. Adler et al. (2010) in ihrer Literaturübersicht zeigen, dass die Kontinuität die Patientenzufriedenheit zwar beeinflusst, aber je nach gewähltem Untersuchungsdesign unterschiedlich starken Einfluss ausübt, was darauf schließen lässt, dass Kontinuität zwar einen beeinflussbaren Patientenzufriedenheitsfaktor darstellt, dieser aber unterschiedlich stark wahrgenommen wird.⁸⁸

Eine ausgedehnte Region im Bereich der Patientenzufriedenheitsforschung nimmt der Faktor Kommunikation ein. Dieser Sachverhalt beruht einerseits, auf den Daten

⁸⁶Vgl. Kim et al. (2004), S. 237.

⁸⁷Vgl. Derksen et al. (2013), S. 76.

⁸⁸Vgl. Adler et al. (2010) (2010), S. 171.

⁸⁵Vgl. Shelton (2000), S. 29.

und Erkenntnissen anderer Fachbereiche, durch welche gezeigt werden konnte, dass Kommunikation einen Zufriedenheitsfaktor darstellt und folglich die positive Arzt-Patienten-Kommunikation auch die Patientenzufriedenheit positiv beeinflussen sollte, andererseits gibt es aber auch eine Vielzahl medizinischer Studien, welche ihren Fokus auf den Zusammenhang einer positiven Arzt-Patienten-Kommunikation und erhöhten Heilungsaussichten bzw. schnellerer Genesung legen.⁸⁹ Eine vielzitierte Studie in diesem Zusammenhang ist jene von Bertakis et al. (1991) aus dem Jahr 1991. Bei dieser Untersuchung wurde die Bedeutung der Arzt-Patienten-Kommunikation definiert und der Zusammenhang zwischen dem medizinischen Gespräch und der Patientenzufriedenheit untersucht. Durch die Aufnahme und Analyse von Arzt-Patienten-Gesprächen konnte festgestellt werden, dass dominante und fachsprachliche Behandlungskommunikation in negativem Zusammenhang mit Patientenzufriedenheit stand, wohingegen partiell private Gesprächsführung die Patientenzufriedenheit positiv beeinflusste. Durch die Studie von Bertakis et al. (1991) konnte ein direkter signifikanter Zusammenhang zwischen dem medizinischen Gespräch und Patientenzufriedenheit nachgewiesen werden.⁹⁰ Bedingt durch die Fülle an Studien und Literatur in Zusammenhang mit Kommunikation und Patientenzufriedenheit wird innerhalb von aktuelleren Studien meist nicht nur die Kommunikation selbst, sondern die Kommunikation, kombiniert mit speziellen Faktoren, untersucht. Ein Beispiel für jene Untersuchungen ist die Studie von Cousin et al., bei welcher der Zusammenhang von Ärztekommunikation und Patienteneinstellung untersucht wurde. Die gesammelten Ergebnisse haben gezeigt, dass fürsorgende Pflegekommunikation bei allen untersuchten Patientengruppen in positivem Zusammenhang mit Patientenzufriedenheit stand, teilnehmende Kommunikation jedoch nur von einigen Patientengruppen als positiv wahrgenommen wurde.⁹¹

Um einen umfassenden Überblick über essentielle, beeinflussbare Patientenzufriedenheitsfaktoren zu erlangen, reicht es jedoch nicht aus, sich mit der Bedeutung und der Positionierung einzelner Faktoren auseinanderzusetzen. Da die Faktorstruktur der Patientenzufriedenheit, wie eingangs erwähnt, ein komplexes, individuell geformtes Gebilde darstellt, ist es unumgänglich, mehrere beeinflussbare Faktoren in die Untersuchung miteinzubeziehen und gegenüberzustellen, um Annahmen über die Konsistenz und Wertigkeit einzelner Zufriedenheitsfaktoren treffen zu können.

Ein Beispiel für die Untersuchung von mehreren beeinflussbaren Faktoren auf die Patientenzufriedenheit ist das Discrete Choice Experiment von Cheraghi-Sohi et al. aus dem Jahr 2008., bei welchem versucht wurde, die Stärke der Patientenprioritäten in Bezug auf die Zufriedenheit mit Hausärztinnen und -ärzten zu messen. Hierfür wurden Patientinnen und Patienten aus sechs englischen Hausarztpraxen gebeten, ihr jeweiliges Prioritätslevel für kürzere Wartezei-

ten, Terminflexibilität, Vertrautheit mit der behandelnden Ärztin / dem behandelnden Arzt, hohe technische Qualität und unterschiedliche Aspekte der Patientenzentriertheit zu bewerten. Die gesammelten Ergebnisse wurden analysiert und jedem Attribut wurde ein monetärer Wert, also die relative Zahlungsbereitschaft für das jeweilige Merkmal, zugeordnet. Den höchsten Wert innerhalb der Untersuchung erreichte die gründliche körperliche Untersuchung (> \$ 40,-), gefolgt von Vertrautheit mit der behandelnden Ärztin / dem behandelnden Arzt (> \$ 10,-) und der Reduktion der Wartezeit (> \$ 5,-). Die Aspekte der Patientenzentriertheit wurden hingegen insgesamt mit einem ähnlich hohen Wert bemessen (ca. \$ 12 - \$ 14). Beeinflusst wurden die Ergebnisse sowohl durch die Art und Dringlichkeit des Arztbesuches als auch durch diverse demografische Aspekte, wie dem Geschlecht, dem Alter oder dem jeweiligen Haushaltseinkommen.⁹²

Eine weitere informative Studie in diesem Zusammenhang ist jene von Paddison et al. aus dem Jahr 2013, bei welcher wiederum untersucht wurde, welche Faktoren die subjektiv empfundene Patientenzufriedenheit am stärksten beeinflussen. Die Grundlage für die diesbezügliche Analyse und Auswertung lieferten die gesammelten Daten der English General Practice Patient Survey 2009/10, an welcher über 2 Millionen Patientinnen und Patienten aus über 8000 Arztpraxen teilnahmen. Die gesammelten Daten lassen darauf schließen, dass die Behandlungskommunikation, gefolgt von der Freundlichkeit des Praxispersonals, den stärksten Einfluss auf die Patientenzufriedenheit ausübt und Zugangsfaktoren, wie die Möglichkeiten der Terminvereinbarung, zwar Einfluss auf die Patientenzufriedenheit ausüben, aber nicht zu den einflussstärksten Faktoren zählen. Die Studie von Paddison et al. unterstützt die Annahme, dass es sich bei Patientenzufriedenheit um ein multidimensionales Konstrukt handelt. Ferner konnte festgestellt werden, dass sowohl die Kommunikation mit der behandelnden Ärztin / dem behandelnden Arzt als auch die Kommunikation mit dem Praxispersonal das ausschlaggebende Qualitätskriterium der kognitiv erfassbaren Behandlungsqualität darstellt.⁹³

Wie die Studien von Cheraghi-Sohi et al. und Paddison et al. zeigen, führen auch Untersuchungen mehrerer beeinflussbarer Patientenzufriedenheitsfaktoren, bedingt durch das gewählte Forschungsdesign und den gewählten Forschungsschwerpunkt, zu unterschiedlichen und nicht oder nur schwer vergleichbaren Ergebnissen. Aus diesem Grund wird in Kapitel 4 – Validierte Skalen und Dimensionen der Patientenzufriedenheit – neben den Grundlagen für die Verwendung vorhandener Skalen und der Beschreibung der Skalen selbst auch dargelegt, welche Zufriedenheitsdimensionen diese messen und aus welchen Faktoren oder Teilfaktoren sich die Dimensionen bzw. Kategorien zusammensetzen.

⁸⁹Vgl. Ha et al. (2010), S. 38.

⁹⁰Vgl. Bertakis et al. (1991), S. 175.

⁹¹Vgl. Cousin et al. (2012), S. 193.

⁹²Vgl. Cheraghi-Sohi et al. (2008), S. 107.

⁹³Vgl. Paddison et al. (2013), S. 1081 ff.

4. Validierte Skalen und Dimensionen der Patientenzufriedenheit

Bevor mit dem Entwurf von selbstentwickelten Erhebungsinstrumenten begonnen wird, sollten vorhandene Studien oder andere Sekundärquellen nach Messinstrumenten durchsucht werden, welche dasselbe oder ein ähnliches Erkenntnisinteresse innehaben wie die angestrebte Untersuchung. Diese Vorgehensweise ist zu empfehlen, da die Verwendung von etablierten und validierten Skalen oder zumindest die Orientierung an solchen im Gegensatz zur Eigenentwicklung und Validierung deutlich weniger personelle und finanzielle Ressourcen in Anspruch nimmt.⁹⁴

Um die Zufriedenheit von Patientinnen und Patienten messbar zu gestalten und etwaige Optimierungspotentiale zu erkennen, wurden in der Vergangenheit zahlreiche Erhebungsinstrumente entwickelt.⁹⁵ Trotz der Vielzahl unterschiedlicher Messskalen sind diese untereinander nur sehr schwer vergleichbar, da viele Forscher unterschiedliche Messansätze verwenden. Außerdem unterscheiden sich die entwickelten bzw. verwendeten Instrumente häufig hinsichtlich ihrer Zielsetzung. Vor diesem Hintergrund wird deutlich, dass die Wahrscheinlichkeit, eine passende vorhandene Skala zu finden abnimmt, je enger die eigene Zielsetzung gefasst ist. Für Fragestellungen mit einem relativ weitläufigen Untersuchungsbereich oder als Grundstein für die Entwicklung von eigenen Erhebungsinstrumenten kann die eingehende Literaturrecherche im Hinblick auf vorhandene Skalen jedoch durchaus zweckdienlich wirken.⁹⁶

Für die Bewertung vorhandener Skalen gelten die Reliabilität und Validität, welche in Kapitel 5 genauer beschrieben sind, als entscheidende Qualitäts- und Gültigkeitskriterien.⁹⁷ Kurzum ist es neben der Validierung des jeweiligen Messinstrumentes, also der Bestimmung, ob auch das gemessen wurde, was gemessen werden sollte, auch notwendig zu überprüfen, ob die durchgeführte Messung reproduzierbar ist.⁹⁸ Dieser Umstand kommt insbesondere zu tragen, wenn für die eigene Untersuchung ein bereits vorhandenes Messinstrument herangezogen wird, da die Erhebungsphase nur praktikabel durchführbar ist, wenn die gewählte vorhandene Skala auch bei Erhebungsdurchführung durch andere Forscher zu vergleichbaren Ergebnissen führt.⁹⁹

Das dritte Beurteilungskriterium hinsichtlich der Qualität vorhandener Messskalen ist die Trennschärfe, welche Auskunft darüber gibt, inwiefern reale Gegebenheiten auch durch das jeweilige Instrument gemessen werden. Außerdem definiert die Trennschärfe das Potential von Messinstrumenten, falsche Annahmen über einen Untersuchungsbereich auch als solche zu erkennen. Grundsätzlich kann hinsichtlich der Anforderungen an vorhandene Skalen zusammengefasst

werden, dass diese den Untersuchungsbereich fokussiert und klar darstellen sollen.¹⁰⁰

Die Literatursuche kann dabei helfen herauszufinden, wie andere Forscher den jeweiligen Untersuchungsbereich abgegrenzt und analysiert haben und führt im Idealfall zu einer gewissen Anzahl an Messinstrumenten, die für die Nutzung in der eigenen Untersuchung verwendet werden können.¹⁰¹ Wenn ein oder mehrere, für die eigene Untersuchung passende, vorhandene Messinstrumente identifiziert worden sind, so bleibt zu entscheiden, ob diese direkt verwendet oder als Fundament zur Entwicklung einer eigenen Messskala herangezogen werden. Als entscheidungsunterstützende Kriterien gelten naturgemäß in erster Linie die Übereinstimmung mit dem eigenen Erkenntnisinteresse sowie die kritische Auseinandersetzung mit der jeweiligen Studie hinsichtlich Reliabilität und Validität.¹⁰² Bei Überlegungen hinsichtlich der Auswahl von vorhandenen Erhebungsinstrumenten sind aber auch die in Tabelle 1 zusammengefassten Grundvoraussetzungen zu beachten.

Zusätzlich zu den in Tabelle 1 beschriebenen prozessualen Grundvoraussetzungen für die Beurteilung vorhandener Messinstrumente bestehen auch rechtliche und monetäre Aspekte, welche bei der Verwendung vorhandener Skalen beachtet werden sollten. Bspw. muss zumindest das schriftliche Einverständnis der Autoren eingeholt werden, wenn das Erhebungsinstrument nicht öffentlich zugänglich ist. Dies gilt ebenso für teilweise öffentlich zugängliche Skalen. Auch wenn ein Instrument in der wissenschaftlichen Literatur veröffentlicht wird, bedeutet dies nicht automatisch, dass es sich im öffentlichen Bereich befindet und keine Genehmigung des Autors und/oder Herausgebers erforderlich ist. Wenn es sich um ein urheberrechtlich geschütztes Instrument handelt, können für den Erwerb oder die Nutzung Gebühren entstehen. Angesichts dessen wird deutlich, dass im Zusammenhang mit Kosteneinsparungspotentialen abgewogen werden muss, ob ein kostenintensives Instrument gekauft, ein kostenloses herangezogen oder ein eigenes Erhebungsinstrument entwickelt wird.¹⁰³

Der Versuch Patientenzufriedenheit messbar zu gestalten, hat in den vergangenen Jahrzehnten viele unterschiedliche Skalen hervorgebracht, wobei der Beginn der wissenschaftlichen Patientenzufriedenheitsmessung auf die siebziger Jahre des zwanzigsten Jahrhunderts zurückgeht.¹⁰⁴ Die ersten Ansätze zur Messung der Patientenzufriedenheit schufen Hulka et al. (1970) mit der Satisfaction with Physician and Primary Care Skala.¹⁰⁵ Danach wurde von Ware und Snyder zur Messung der Zufriedenheit mit medizinischen Dienstleistungen die erste Version des Patient Satisfaction Questionnaire entwickelt.¹⁰⁶ Anschließend konzipierten Larsen et al. (1979)

¹⁰⁰Vgl. Raab et al. (2009), S. 68 f.

¹⁰¹Vgl. Kimberlin und Winterstein (2008), S. 2280.

¹⁰²Vgl. Steiner et al. (2015), S. 8.

¹⁰³Vgl. Kimberlin und Winterstein (2008), S. 2280 f.

¹⁰⁴Vgl. Gill und White (2009), S. 11.

¹⁰⁵Vgl. Hulka et al. (1970), S. 429 ff, zitiert nach Gill und White (2009), S. 11.

¹⁰⁶Vgl. Ware und Snyder (1975), S. 669 ff, zitiert nach Gill und White

⁹⁴Vgl. Kimberlin und Winterstein (2008), S. 2280.

⁹⁵Vgl. Thayaparan und Mahdi (2013), S. 1.

⁹⁶Vgl. Steiner et al. (2015), S. 7.

⁹⁷Vgl. Kimberlin und Winterstein (2008), S. 2276.

⁹⁸Vgl. Homburg und Krohmer (2006), S. 255.

⁹⁹Vgl. Steiner et al. (2015), S. 9.

Tabelle 1: Grundvoraussetzungen für die Beurteilung vorhandener Messinstrumente; Quelle: In Anlehnung an Kimberlin und Winterstein (2008), S. 2280 f.

Voraussetzung	Beschreibung
Erkenntnisinteresse definieren	Bevor mit der Suche nach vorhandenen Skalen begonnen wird, sollte das eigene Erkenntnisinteresse bereits möglichst genau definiert und das Untersuchungsdesign bereits festgelegt bzw. abgegrenzt worden sein, um die jeweilige Passgenauigkeit einschätzen zu können.
Datengenerierung überprüfen	Um zu beurteilen, ob die jeweilige Skala zum eigenen Forschungsgebiet passt, reicht es nicht aus, sich nur mit dem Titel oder der Zusammenfassung der jeweiligen Arbeit auseinanderzusetzen. In der Regel kann die Beurteilung der Passgenauigkeit auch nicht direkt von den einflussnehmenden bzw. untersuchten Variablen abgeleitet werden. Aus diesen Gründen muss das in Frage kommende Instrument hinsichtlich der Fokussierung der Fragen, vor allem aber hinsichtlich der Datengenerierung, überprüft werden, um beurteilen zu können, ob dieses für die Unterstützung oder Durchführung der eigenen Erhebung in Frage kommt.
Gütekriterien untersuchen	Neben der Überprüfung der Gütekriterien Validität, Reliabilität und gegebenenfalls auch Trennschärfe muss untersucht werden, ob die jeweilige Skala für eine Population entwickelt wurde, die annähernd der zu untersuchenden Population entspricht.
Fehlende bzw. fehlerhafte Daten hinterfragen	Die Studien sind hinsichtlich fehlender oder fehlerhafter Daten (z.B. Antwortkategorien, die häufig nicht beantwortet wurden) zu überprüfen und kritisch zu hinterfragen.
Untersuchungsdesign bzgl. Passgenauigkeit überprüfen	Wird eine Skala übernommen, so muss darauf geachtet werden, dass das verwendete Untersuchungsdesign und -umfeld dem angestrebten entspricht. So ist es bspw. nicht sinnvoll, ein Instrument, welches für die mündliche Befragung entwickelt wurde, als Onlinebefragungsinstrument zu verwenden, da etwaige Erläuterungen nicht durchgeführt werden können.

den Client Satisfaction Questionnaire zur Beurteilung der allgemeinen Patientenzufriedenheit mit gesundheitlichen Dienstleistungen, welcher jedoch 1984 durch die Patient Satisfaction Skala abgelöst wurde.¹⁰⁷ Seit den Anfängen der Patientenzufriedenheitsmessung wurden in diesem Bereich zahlreiche Messinstrumente und Skalen konstruiert, wobei diese hinsichtlich Validität und Reliabilität durchaus variieren und die Qualität nicht immer beurteilt werden kann.¹⁰⁸

Um eine detaillierte Übersicht über unterschiedliche Skalen und Dimensionen der Patientenzufriedenheit zu geben, werden in den folgenden Kapiteln, 4.1 und 4.2, fünf internationale bzw. multinationale und zwei deutschsprachige Skalen zur Messung der Patientenzufriedenheit mit Hausärztinnen und -ärzten vorgestellt. Im Resümee werden die gesammelten Ergebnisse gebündelt dargestellt. Ausgewählte Illustrationen der untersuchten Skalen zur Messung der Patientenzufriedenheit sind im Anhang ab Seite 108 zu finden.

4.1. Internationale Skalen zur Messung von Patientenzufriedenheit

Im Folgenden werden fünf internationale bzw. multinationale Skalen zur Messung der Patientenzufriedenheit vorgestellt. Um eine umfassende Übersicht in diesem Bereich zu

(2009), S. 11.

¹⁰⁷Vgl. Larsen et al. (1979), S. 197 ff, zitiert nach Gill und White (2009), S. 11.

¹⁰⁸Vgl. Gill und White (2009), S. 11.

erstellen, wurde pro Dekade von 1980 bis 2017 jeweils eine Patientenzufriedenheitserhebung ausgewählt, in welcher entweder selbst eine Patientenzufriedenheitsskala entwickelt oder eine etablierte Skala verwendet wurde. Das zweite Kriterium, um in die Betrachtung mitaufgenommen zu werden, war, dass innerhalb der jeweiligen Untersuchung die erhobenen Zufriedenheitsdimensionen oder Kategorien definiert sein mussten. Um Aktualität zu gewährleisten, wurde auch eine Untersuchung aus dem Jahr 2017 ausgewählt. Durch die gewählte Such- und Auswahlstrategie war es also nicht zwingend notwendig, dass Skala und Erhebung im selben Jahr veröffentlicht wurden. Dieser Umstand kommt insbesondere bei der untersuchten internationalen Skala zu tragen. Hier wurde die Erhebung zwar 2017 durchgeführt, jedoch mit einer Skala, die bereits 1994 entwickelt wurde. Details zur Such- und Auswahlstrategie sind im Anhang ab Seite 101 zu finden.

Die durch Einsatz der Suchstrategie identifizierten und evaluierten internationalen Skalen zur Messung der Patientenzufriedenheit sind der Patient Satisfaction Questionnaire Form 2 (PSQF2), die Patientenzufriedenheitsskala der European Task Force on Patient Evaluations of General Practice (EUROPEP), die General Practice Assessment Survey (GPAS), der Questionnaire of Patient Satisfaction und die Weiterentwicklung des Patient Satisfaction Questionnaires, der Patient Satisfaction Questionnaire Short Form (PSQ 18). In Tabelle 2 sind die ausgewählten Skalen in Hinblick auf Fragen

bzw. Item-Anzahl, Antwortskalenformat, Veröffentlichungsjahr, Entwicklungs- und Erstverwendungsort und Anzahl an gemessenen Zufriedenheitskategorien bzw. Dimensionen einander gegenübergestellt.

4.1.1. Patient Satisfaction Questionnaire Form 2 (PSQF2)

Um das Konstrukt der Patientenzufriedenheit messbar zu gestalten, entwickelten Ware et al. an der Southern Illinois University School ein Messinstrument namens PSQ (Patient Satisfaction Questionnaire), wobei die Forscher Anfang 1983 eine Überarbeitung der Skala mit dem Namen Patient Satisfaction Questionnaire Form 2 (PSQF2) veröffentlichten. Das Hauptziel des Projektes war es, ein relativ simples Befragungsinstrument zur Messung von Zufriedenheit mit ärztlichen und institutionellen Leistungen zu konzipieren. Demnach wurde der PSQ entwickelt, um die Zufriedenheit mit niedergelassenen und stationären Ärztinnen und Ärzten, aber auch mit Arztpraxen, Krankenhäusern sowie regionalen und politischen Rahmenbedingungen, zu messen. Der Fragebogen besteht aus 55 Punkten, welche jeweils auf einer fünfstufigen Likert-Skala zu bewerten sind und misst sowohl die Patientenzufriedenheit mit zentralen Eigenschaften von Ärztinnen und Ärzten als auch die generelle Gesamtzufriedenheit mit der zugänglichen medizinischen Versorgung.¹⁰⁹ Die Likert-Skala wurde als Antwortformat, von starker bis hin zu schwacher Zustimmung, gewählt, da diese bereits in vorangegangenen Untersuchungen vorteilhaft eingesetzt wurde, wobei dadurch auch für die Patientinnen und Patienten ein relativ einfacher Beantwortungsprozess gewährleistet werden konnte.¹¹⁰ Einen weiteren Vorteil bei Verwendung dieser Antwortskala stellten die kalkulatorischen Möglichkeiten dar.¹¹¹ So wurden den unterschiedlichen Antwortkategorien Werte von 0 (überhaupt keine Zustimmung) bis 100 (starke Zustimmung) zugeordnet, um die gesammelten Daten in der Auswertungsphase berechenbar zu gestalten.¹¹²

Durch die häufige Verwendung und Überarbeitung ist der PSQF2 zu einer etablierten Befragungsskala im Bereich der Patientenzufriedenheitsmessung avanciert. Trotzdem bestehen einige Kritikpunkte, die sich hauptsächlich auf die in den Fragen enthaltenen Konzeptdefinitionen und die Anwendungsgebiete der Skala beziehen. So wird bspw. die empfundene Dienstleistungsqualität gemessen, diese wird jedoch nicht definiert oder vereinheitlicht, was die Vergleichbarkeit von, mit dem persönlichen Verständnis verknüpften, Empfindungen erschwert. Auch hinsichtlich der internationalen Verwendung der Skala können Herausforderungen auftreten, da Qualitätsmerkmale unterschiedlich wahrgenommen werden und auch bei der Untersuchung unterschiedlicher Länder und / oder Kulturen voneinander abweichen können.¹¹³ Trotz der genannten Kritikpunkte haben Untersuchungen gezeigt, dass der PSQF2 zu zufriedenstellenden Ergebnissen hinsichtlich

der Gütekriterien führt und die Skala häufig genannte Zufriedenheitsdimensionen zuverlässig misst.¹¹⁴

Mit dem PSQF2 werden unter anderem die Zufriedenheit mit der empfundenen Servicequalität, der wahrgenommenen technischen Qualität und den regionalen Rahmenbedingungen¹¹⁵ auf einer multidimensionalen Skala gemessen.¹¹⁶ Im Bereich der Zufriedenheitsdimensionen unterscheiden Ware et al. insgesamt acht Faktorkategorien, welche die Patientenzufriedenheit beeinflussen. Diese sind zwischenmenschliche Belange, technische Qualität, Erreichbarkeit, finanzielle Faktoren, Effizienz, Kontinuität, physische Gegebenheiten und Verfügbarkeit. Unter zwischenmenschlichen Belangen wird innerhalb der Untersuchung der persönliche Umgang mit den Patientinnen und Patienten, Freundlichkeit und Höflichkeit verstanden. Die technische Qualität beschreibt die fachliche Kompetenz, Gründlichkeit, Genauigkeit sowie die Vermeidung von Risiken und Fehlern. Erreichbarkeit beinhaltet sowohl den Aufwand, welcher betrieben werden muss, um einen Termin zu erhalten als auch die Wartezeit auf den Termin bzw. die Wartezeit in der Praxis sowie die Erreichbarkeit in Bezug auf die Lage der Ordination. Finanzielle Faktoren beinhalten die direkt der Behandlung zurechenbaren Kosten und die Kosten des Versicherungsschutzes. Die Effizienz beurteilt das Ergebnis der medizinischen Versorgung sowie die Unterstützung bei der Verbesserung und Erhaltung der Gesundheit. Unter Kontinuität wird einerseits die Beständigkeit der geografischen Lage der Arztpraxis, andererseits die Beständigkeit der Ärztin / des Arztes selbst, bspw. in Gemeinschaftspraxen oder Krankenhäusern, verstanden. Physische Gegebenheiten beziehen sich auf die ordnungsgemäße Ausstattung und eine angenehme Atmosphäre und Verfügbarkeit summiert die Arzt- bzw. Krankenhauswahlmöglichkeiten.¹¹⁷ Die unterschiedlichen Faktorkategorien sind in Tabelle 3 charakterisiert und übersichtlich gegenübergestellt.

Die Studie von Ware et al. findet auch heute noch, insbesondere in Bezug auf die Kategorisierung von Zufriedenheitsdimensionen, beachtlichen Zuspruch und wird häufig als Basis zur Entwicklung von Instrumenten zur Messung von Patientenzufriedenheit herangezogen. Dieser Umstand ist unter anderem darauf zurückzuführen, dass der PSQF2 allgemeine Dimensionen für eine Vielzahl an unterschiedlichen Anwendungsbereichen beinhaltet.¹¹⁸

4.1.2. European Task Force on Patient Evaluations of General Practice (EUROPEP)

Das Erhebungsinstrument der European Task Force on Patient Evaluations of General Practice (EUROPEP) ist ein 23 Fragen umfassender Fragebogen, welcher verschiedene Zufriedenheitsaspekte mit der hausärztlichen Versorgung, unter Miteinbezug von fünfstufigen Likert-Skalen misst, wobei

¹⁰⁹Vgl. Ware et al. (1983), S. 247.

¹¹⁰Vgl. Baker (1990), S. 487.

¹¹¹Vgl. Wüthrich-Schneider (2000b), S. 1118.

¹¹²Vgl. Hughes et al. (2000), S. 2878.

¹¹³Vgl. Aldana et al. (2001), S. 516.

¹¹⁴Vgl. Ware et al. (1983), S. 247.

¹¹⁵Vgl. Dagger et al. (2007), S. 125.

¹¹⁶Vgl. Sixma et al. (1998), S. 86.

¹¹⁷Vgl. Ware et al. (1983), S. 248.

¹¹⁸Vgl. Jalil et al. (2017), S. 2 f. und die dort zitierte Literatur und Hagedoorn et al. (2003), S. 254 ff.

Tabelle 2: Ausgewählte internationale Skalen zur Patientenzufriedenheitserhebung; Quelle: Verfasser.

Skala	Fragen	Skalenformat	Veröffentlicht	Entwickelt in	Dimensionen / Überkategorien
PSQF2	55	5-stufig / Likert	1984	USA	8
EUROPEP	23	5-stufig / Likert	1996	EU	5
GPAS	53	Verschiedene	2000	GB	10
QPS	20	Verschiedene	2012	SRB	11
PSQ 18	18	5-stufig / Likert	1994	USA	7

Tabelle 3: Patientenzufriedenheitsdimensionen nach Ware et. al.; Quelle: In Anlehnung an Ware et al. (1983), S. 248.

Faktorkategorie	Inhaltliche Definition
Zwischenmenschliche Belange	Persönlicher Umgang mit der Patientin / dem Patienten
Technische Qualität	Behandlungs- und Diagnosekompetenz
Erreichbarkeit	Organisation der medizinischen Versorgung
Finanzielle Faktoren	Monetäre Belastungen
Effizienz	Behandlungsergebnis
Kontinuität	Gleichbleiben der Anbieterin /des Anbieters
Physische Gegebenheiten	Ausstattungsmerkmale
Verfügbarkeit	Präsenz medizinischer Ressourcen

die Antwortkategorien von „schlecht“ bis „ausgezeichnet“ reichen.¹¹⁹ Dieses Instrument wurde entwickelt, um die empfundene Patientenzufriedenheit mit hausärztlicher Pflege international vergleichbar zu gestalten und dadurch den behandelnden Hausärztinnen und -ärzten die Möglichkeit zu geben, die Qualität der Primärversorgung zu verbessern.¹²⁰

Das Befragungsinstrument wurde als Untersuchungsskala unterschiedlicher gleichwertiger Zufriedenheitsdimensionen konzipiert, wobei der Entwicklungsprozess mit Hilfe von Literaturrecherche, Experteninterviews sowie Prioritäts- und Validierungsstudien realisiert wurde.¹²¹ Im Gegensatz zu anderen Instrumenten zur Patientenzufriedenheitsmessung wird mit Hilfe der EUROPEP-Skala nicht versucht, eine punktuelle Bestandsaufnahme, bspw. bezogen auf die Zufriedenheit mit der momentanen Behandlung, durchzuführen. Viel mehr liegt der Fokus darauf, die empfundene Zufriedenheit während der vergangenen Monate bis hin zu einem Jahr zu messen, wobei der Fragebogen in der Regel entweder selbstständig und nach Behandlungsende durch die Patientinnen und Patienten ausgefüllt oder nach Zusendung und Unterfertigung von der jeweiligen Patientin / vom jeweiligen Patienten per E- Mail retourniert wird.¹²²

Durch die weite Verbreitung auch außerhalb Europas, hat sich die EUROPEP-Skala zu einem weithin bekannten

und vielgenutzten Instrument zur Erhebung der Zufriedenheit mit Hausarztpraxen aus Patientenperspektive entwickelt.¹²³ Doch die große Stärke – die Möglichkeit, Patientenzufriedenheit in unterschiedlichen politischen und kulturellen Regionen zu untersuchen – scheint auch einige Schwächen zu beinhalten. So konnte bspw. gezeigt werden, dass sich die innerhalb des Befragungsinstrumentes verwendete Faktorstruktur nicht auf alle Kulturkreise übertragen lässt, was innerhalb dieser Bereiche zu niedriger Ergebnisvalidität führt.¹²⁴ Dieser Umstand lässt darauf schließen, dass das EUROPEP- Befragungsinstrument für die Anwendung in bestimmten Ländermärkten adaptiert bzw. angepasst werden müsste.¹²⁵ Auch aktuellere Studien, wie jene von Dimova et al., welche zur Validierung der bulgarischen EUROPEP-Skala durchgeführt wurde, kommen zu ähnlichen Ergebnissen. Die statistische Datenauswertung lässt darauf schließen, dass zumindest die Vergleichbarkeit zwischen den Zufriedenheitsdimensionen einzelner Patientinnen und Patienten verfälscht ist.¹²⁶ Andere Untersuchungen, wie jene zur Validierung des italienischen EUROPEP-Befragungsinstrumentes, konnten hingegen bestätigen, dass die EUROPEP-Skala in unterschiedlichen Länderversionen ein valides und einfach zu handhabendes Instrument ist, um Informationen über die Patientenerfahrung und die Patientenzufriedenheit in

¹¹⁹Vgl. Wensing et al. (2002), S. 112.

¹²⁰Vgl. Grol et al. (2000), S. 883.

¹²¹Vgl. Vedsted et al. (2008), S. 175.

¹²²Vgl. Hogan et al. (2017), S. 173.

¹²³Vgl. Bjertnaes et al. (2011), S. 342.

¹²⁴Vgl. Vedsted et al. (2008), S. 179.

¹²⁵Vgl. Bjertnaes et al. (2011), S. 342.

¹²⁶Vgl. Dimova et al. (2017), S. 72.

und mit Hausarztpraxen zu untersuchen.¹²⁷ Obwohl das EUROPEP-Erhebungsinstrument bereits Ende des vergangenen Millenniums entwickelt wurde, spielt es im Bereich der Patientenzufriedenheitsmessung immer noch eine erhebliche Rolle¹²⁸ und wird auch heute noch zur Untersuchung mehrerer oder einzelner Zufriedenheitsfaktoren in Bezug auf die hausärztliche Versorgung angewandt.¹²⁹

Die EUROPEP-Skala untersucht sowohl organisationale als auch arzt spezifische Patientenzufriedenheitsfaktoren.¹³⁰ Grol et al. gruppieren die mit dem EUROPEP-Erhebungsinstrument untersuchten Zufriedenheitsfaktoren innerhalb der fünf Kategorien medizinisch-technische Betreuung, Arzt-Patienten-Beziehung, Auskunft und Hilfestellung, Verfügbarkeit und Erreichbarkeit sowie Organisation. Im Gegensatz zur medizinisch-technischen Betreuung, welche die fachliche Kompetenz beurteilt, bezieht sich die Arzt-Patienten-Beziehung eher auf emotionale Kompetenzen, Empathie und Sympathie der behandelnden Ärztinnen und Ärzte. Die Auskunft und Hilfestellung bezieht sich auf die fachliche Information und Kommunikationsbereitschaft und mittels Verfügbarkeit und Erreichbarkeit werden alle Aspekte summiert, welche mit einer bestimmten Dauer zusammenhängen, wie Wartezeit auf einen Termin oder die Konsultationsdauer. Abschließend beurteilt die organisationale Dimension Aspekte, wie die physische Praxisbeschaffenheit, aber auch die Zusammenarbeit mit Praxismitarbeiterinnen und -mitarbeitern.¹³¹ Tabelle 4 fasst die unterschiedlichen Faktorkategorien übersichtlich zusammen.

Die EUROPEP-Skala wird häufig herangezogen, um grundlegende Kategorien für die Messung von Patientenzufriedenheit innerhalb unterschiedlicher geografischer Lagen zu bilden.¹³² Aber auch innerhalb der Analysen einzelner Länder wird die verwendete Kategorisierung in fünf Zufriedenheitsdimensionen häufig angewandt.¹³³ Generell hat sich das EUROPEP-Befragungsinstrument zu einem viel diskutierten und häufig validierten Erhebungsinstrument entwickelt, welches häufig auch in seinen unterschiedlichen nationalen Ausprägungen Bestand hat.¹³⁴

4.1.3. General Practice Assessment Survey (GPAS)

Die General Practice Assessment Survey (GPAS) ist ein Instrument zur Erhebung der Patientenzufriedenheit mit Hausarztpraxen. Das Zufriedenheitsniveau setzt sich hierbei aus neun Zufriedenheitsdimensionen zusammen und wird mit Hilfe von unterschiedlichen Antwortskalen erhoben.¹³⁵ Der Großteil der Untersuchungselemente wird jedoch mit Hilfe einer sechsstufigen Skala gemessen, welche von „sehr schlecht“ bis „ausgezeichnet“ reicht. Gewichtet werden die

unterschiedlichen Antwortkategorien als Prozentsatz zwischen 0 und 100, wobei „sehr schlecht“ dem Prozentsatz 0 und „ausgezeichnet“ dem Prozentsatz 100 entspricht.¹³⁶ Der Erhebungsbereich, welcher nicht dem Skalenniveau der übrigen Erhebungskategorien entspricht, beschäftigt sich mit dem Gesamtvertrauen in die behandelnde Hausärztin / den behandelnden Hausarzt. Die Vertrauenskategorie wird auf einer Skala von 1 bis 10 gemessen, wobei die Antwort 1 keinerlei Vertrauen und die Antwort 10 vollkommenem Vertrauen entspricht. Die verbleibenden Antwortkategorien messen unter anderem die wahrgenommene Qualität der Arzt-Patienten-Beziehung und Kontinuitätsaspekte.¹³⁷

Bei der GPAS handelt es sich um ein reliables, validiertes Erhebungsinstrument, welches in vielerlei Hinsicht eingesetzt werden kann. Durch die Definition der Zufriedenheitsdimensionen und den klaren Fokus auf Hausarztpraxen eignet sich die GPAS exzeptionell als unterstützendes Zufriedenheitsanalyseinstrument für einzelne oder zusammengeslossene Hausarztpraxen sowie Forscher, welche die primären Zufriedenheitsareale hausärztlicher Versorgung untersuchen.¹³⁸

Obwohl die GPAS auf breite Akzeptanz trifft, werden in der Literatur einige Mängel hinsichtlich der Datenerhebung innerhalb der Primäruntersuchung postuliert. So bestehen bspw. signifikante Messwertdifferenzen zwischen Erhebungen, welche innerhalb der Hausarztpraxis durchgeführt wurden und jenen, die per Post erhoben wurden, wobei nicht geklärt ist, ob diese Unterschiede auf die jeweilige Erhebungsmethode oder auf die Patientinnen und Patienten selbst zurückzuführen sind.¹³⁹ Des Weiteren wurde die GPAS generell nicht entworfen, um Standards in Bezug auf die Kategorisierung von Patientenzufriedenheit festzulegen, sondern um diese messbar zu gestalten. Aus diesem Grund muss vor Verwendung bzw. Anwendung der GPAS überprüft werden, ob die vorgegebenen Patientenzufriedenheitsdimensionen auch dem eigenen Untersuchungsrahmen entsprechen.¹⁴⁰ Trotz der genannten Mängel innerhalb der Primärstudie hat sich die GPAS zu einem wichtigen Instrument zur Bewertung mehrerer Zufriedenheitsdimensionen mit der Primärversorgung entwickelt.¹⁴¹

Campbell et al. zogen die bereits 1999 durch das National Primary Care Research and Development Centre in Manchester entwickelte GPAS-Skala heran, um den Zusammenhang von Alter, Geschlecht und Ethnie von Patientinnen und Patienten und deren Beurteilung von Hausärztinnen und -ärzten zu untersuchen.¹⁴² Zu den einflussnehmenden Patientenzufriedenheitsfaktorkategorien zählten hierbei der Zugang zu medizinischer Versorgung, das Pflegeniveau und das gesammelte Wissen der Ärztinnen und Ärzte über

¹²⁷Vgl. Milano et al. (2007), S. 92.

¹²⁸Vgl. Yaphe (2015), S. 162.

¹²⁹Vgl. Sebo et al. (2016), 120.

¹³⁰Vgl. Hogan et al. (2017), S. 173.

¹³¹Vgl. Grol et al. (1999), S. 5.

¹³²Vgl. Schäfer et al. (2015), S. 161.

¹³³Vgl. Sebo et al. (2015), S. 2.

¹³⁴Vgl. Roque et al. (2015), S. 5 f.

¹³⁵Vgl. Ramsay et al. (2000), S. 372.

¹³⁶Vgl. Bower und Roland (2002), S. 126.

¹³⁷Vgl. Tarrant et al. (2003), S. 798.

¹³⁸Vgl. Ramsay et al. (2000), S. 372.

¹³⁹Vgl. Bower und Roland (2002), S. 126 ff.

¹⁴⁰Vgl. Bower et al. (2002), S. 3.

¹⁴¹Vgl. Ramsay et al. (2000), S. 372.

¹⁴²Vgl. Campbell et al. (2001), S. 91.

Tabelle 4: Patientenzufriedenheitsdimensionen nach Grol et al.; Quelle: In Anlehnung an Grol et al. (1999), S. 4 ff.

Faktorkategorie	Inhaltliche Definition
Medizinisch-technische Betreuung	Fachliche Kompetenz
Arzt-Patienten-Beziehung	Emotionale Kompetenz
Auskunft und Hilfestellung	Informations- und Kommunikationsbereitschaft
Verfügbarkeit und Erreichbarkeit	Zeitliche Aspekte
Organisation	Ausstattungs- und Organisationsmerkmale

ihre Patientinnen und Patienten,¹⁴³ sowie technische Aspekte, Kommunikation, zwischenmenschliche Belange, Vertrauen, das Praxispersonal und die Kontinuität der behandelnden Ärztin / des behandelnden Arztes und der gesamte Behandlungsprozess.¹⁴⁴ Zugang zur medizinischen Versorgung beinhaltet hierbei Aspekte, wie die Lage und Öffnungszeiten der Praxis, die telefonische Erreichbarkeit und die Wartezeit auf einen Termin oder im Warteraum. Der Faktor Praxispersonal misst die wahrgenommene personenbezogene Servicequalität, welche nicht die jeweilige Ärztin / den jeweiligen Arzt betrifft. Kontinuität beschreibt die langfristige Pflege durch dieselbe Hausärztin / denselben Hausarzt. Zwischenmenschliche Belange summieren persönliche Eigenschaften wie Geduld, Fürsorge und Betroffenheit und der Faktor Kommunikation bewertet die kommunikativen Fähigkeiten der Ärztin / des Arztes. Technische Aspekte bewerten sowohl das medizinische Gesamtwissen als auch die praktische Anwendung und Vertrauen beschreibt die der jeweiligen Ärztin / dem jeweiligen Arzt entgegengebrachte Zuversicht. Die Patientenkenntnis bezieht sich auf das Wissen, welches über persönliche Aspekte der Patientin / des Patienten vorhanden ist und die Pflegedimension misst die durch das Pflegepersonal, bspw. Krankenpfleger oder medizinische Assistentinnen und Assistenten, wahrgenommene Servicequalität. Abschließend wird mit dem Faktor Behandlungsprozess die Gesamtzufriedenheit mit der hausärztlichen Versorgung gemessen.¹⁴⁵ Eine Übersicht über die unterschiedlichen Faktorkategorien bietet Tabelle 5.

4.1.4. Questionnaire of Patient Satisfaction (QPS)

Der Questionnaire of Patient Satisfaction (QPS) ist ein 20 Fragen umfassendes Instrument zur Erhebung der wahrgenommenen qualitativen Behandlungsdimension, der Behandlungsstruktur, dem Behandlungsprozess und dem Behandlungsergebnis sowie dem daraus resultierenden Patientenzufriedenheitslevel.¹⁴⁶ Das Instrument wurde explizit für die Untersuchung der wahrgenommenen Servicequalität und der daraus resultierenden Zufriedenheit mit Hausarztpraxen entwickelt, wobei einzelne Kategorien mit unterschiedlichen Frageblöcken bedacht wurden. So beziehen sich die ersten

acht Fragen ausschließlich auf demografische und sozioökonomische Aspekte und die restlichen zwölf auf die Patientenzufriedenheit.¹⁴⁷

Der QPS wurde von Vukovic et al. entwickelt, um die Gültigkeit einzelner festgelegter Zufriedenheitsdimensionen und die Zufriedenheit mit den allgemeinen Praxisleistungen zu messen. Die Studie, an welcher insgesamt 1314 Patientinnen und Patienten teilnahmen, wurde innerhalb einer Zeitspanne von zwei aufeinanderfolgenden Jahren im General Medicine Department of Health Center Valjevo in Serbien durchgeführt.¹⁴⁸

Innerhalb des QPS definierten Vukovic et al. die zehn Patientenzufriedenheitsfaktorkategorien Sympathie, Fachwissen, Behandlungszeit, Information, Wartezeit, Erreichbarkeit, Respekt, Ernsthaftigkeit, zwischenmenschliche Beziehung und Zufriedenheit mit dem Praxispersonal. Sympathie beschreibt hierbei den Grad an empfundener Sympathie, Fachwissen den Grad an wahrgenommener fachlicher Kompetenz der behandelnden Ärztin / des behandelnden Arztes. Des Weiteren stellt die Dimension Behandlungszeit dar, ob der Patientin / dem Patienten den Umständen entsprechend ausreichend Untersuchungs- und Gesprächszeit entgegengebracht wurde und Information beschreibt in diesem Zusammenhang die ausreichende Wissensvermittlung bzgl. des Krankheitsbildes, auftretenden Symptomen, Behandlungsmöglichkeiten und möglichen Nebenwirkungen. Die Wartezeit definiert als zweite temporale Dimension die Zeitspanne von der Anmeldung bis zum Konsultationsbeginn und die Erreichbarkeit die Möglichkeiten der Kontaktaufnahme mit der behandelnden Ärztin / dem behandelnden Arzt. Die beiden Dimensionen Respekt und Ernsthaftigkeit messen, inwiefern die Patientin / der Patient als eigenständiges Individuum wahrgenommen wird, wobei Respekt beschreibt, inwieweit die Patientenmeinung in die Behandlung miteinfließt und Ernsthaftigkeit, wie viel Vertrauen der Person und den geschilderten Symptomen entgegengebracht wird. Die zwischenmenschliche Beziehungsdimension beschreibt die persönlichen Eigenschaften der behandelnden Ärztin / des behandelnden Arztes und die zehnte Zufriedenheitsdimension beschreibt die generelle Zufriedenheit mit dem Praxispersonal. Abschließend wird durch die Untersuchung der Gesamtzufriedenheit die generelle Zufriedenheit mit allen

¹⁴³Vgl. Bower et al. (2002), S. 489.

¹⁴⁴Vgl. Tarrant et al. (2003), S. 798.

¹⁴⁵Vgl. Campbell et al. (2001), S. 91.

¹⁴⁶Vgl. Ogaji (2017), S. 24.

¹⁴⁷Vgl. Vukovic et al. (2012), S. 711.

¹⁴⁸Vgl. Janicijevic et al. (2013), S. 161.

Tabelle 5: Patientenzufriedenheitsdimensionen nach Campbell et al.; Quelle: In Anlehnung an Campbell et al. (2001), S. 91.

Faktorkategorie	Inhaltliche Definition
Zugang zu med. Versorgung	Erreichbarkeit und Wartezeit
Technische Aspekte	Fachliche Kompetenz
Kommunikation	Kommunikationsfähigkeit
Zwischenmenschliche Belange	Persönliche Eigenschaften der Ärztin / des Arztes
Vertrauen	Zuversicht in Diagnose und Behandlung
Patientenkenntnis	Wissen über persönliche Aspekte der Patientinnen / Patienten
Pflege	Wahrgenommene Servicequalität des Pflegepersonals
Personal	Wahrgenommene Servicequalität des Personals
Kontinuität	Gleichbleiben der Anbieterin / des Anbieters
Behandlungsprozess	Gesamtzufriedenheit mit hausärztlicher Versorgung

untersuchten Merkmalen betrachtet.¹⁴⁹ Die gesammelten Faktorkategorien sind in Tabelle 6 dargestellt.

In der Studie von Vukovic et al. konnte klar dargestellt werden, dass Patientenzufriedenheit einerseits durch das Niveau der Patientenzentriertheit des medizinischen Personals, andererseits aber auch durch die wahrgenommene Aktualität der Gesundheitsdienstleistung beeinflusst wird.¹⁵⁰ Insbesondere wird die Untersuchung von Vukovic et al. herangezogen, um temporale Aspekte der Patientenzufriedenheit zu verdeutlichen.¹⁵¹

4.1.5. Patient Satisfaction Questionnaire Short Form (PSQ 18)

Umfragen, welche von den befragten Personen als lang und zeitaufwendig wahrgenommen werden, liefern zwar in der Regel mehr Daten bzw. Datenpunkte, können aber hinsichtlich der Erhebungsquote und der Datenqualität negative Auswirkungen zeigen. Vor diesem Hintergrund wurden auch im Bereich der Patientenzufriedenheitsbefragung Kurzformversionen etablierter Messskalen geschaffen. Ein Beispiel von solch gekürzten bzw. auf das Wesentliche reduzierten Skalen ist der Patient Satisfaction Questionnaire Short Form (PSQ 18).¹⁵² Dieses Erhebungsinstrument wurde für die Untersuchung der Patientenzufriedenheit mit niedergelassenen und klinischen Ärztinnen und Ärzten entwickelt und besteht aus 18 Fragen, welche jeweils auf einer fünfstufigen Likert-Skala zu beantworten sind. Die mehrdimensionale Skala wurde von Marshall und Hays bereits 1994 entwickelt, um die Zufriedenheit von Patientinnen und Patienten direkt nach dem Arztbesuch zu messen.¹⁵³

Der PSQ 18 ist eine auf die wesentlichen Merkmalsausprägungen reduzierte Version des 50 Fragen umfassenden Patient Satisfaction Questionnaire Form 3 (PSQF3). Der

PSQF3 ist wiederum eine Weiterentwicklung, mit Zusammenlegung und Variation von Zufriedenheitsfaktorkategorien, des PSQF2. Trotz der Fragenreduktion wird beim PSQ 18 keine der sieben Zufriedenheitsdimensionen des PSQF3 ausgeschlossen. Darüber hinaus konnte durch Überprüfung der gesammelten Daten eine direkte Beziehung zwischen dem Antwortverhalten von Befragungen mittels dem PSQF3 und dem PSQ 18 dargestellt werden.¹⁵⁴ Der PSQ 18 ist ein validiertes und reproduzierbares Zufriedenheitserhebungsinstrument, welches einerseits durch seine vielfältigen Einsatzmöglichkeiten besticht, andererseits aber auch durch seinen geringen Fragenumfang von Patientinnen und Patienten sehr gut aufgenommen wird und sich dadurch ideal für kurze Untersuchungen in Bezug auf die Patientenzufriedenheit eignet.¹⁵⁵

Der PSQ 18 beinhaltet insgesamt 18 Patientenzufriedenheitseigenschaften, welche innerhalb der sieben Faktorkategorien Kommunikation, zwischenmenschliche Belange, Erreichbarkeit und Komfort, technische Qualität, Konsultationsdauer, finanzielle Aspekte und generelle Zufriedenheit gruppiert sind.¹⁵⁶ Unter Kommunikation wird hierbei sowohl die Kommunikationsfähigkeit als auch die Fähigkeit der Patientin / dem Patienten zuzuhören und auf ihre / seine Wünsche einzugehen verstanden. Ferner beschreibt die Konsultationsdauer die aufgewendete Behandlungszeit sowie die entspannte Atmosphäre und zwischenmenschliche Belange, betonen den höflichen und freundlichen Umgang miteinander. Im Gegensatz dazu summieren die Faktorkategorien Erreichbarkeit und Komfort Aspekte wie den Zugang zur Erstversorgung, die Wartezeit und die Möglichkeit zur Konsultation der gewünschten Ärztin / des gewünschten Arztes. Die technische Qualität bezieht sich auf die Praxisausstattung und die fachliche Kompetenz sowie auf die wahrgenommene Fürsorge und mögliche Zweifel in Bezug

¹⁴⁹Vgl. Vukovic et al. (2012), S. 712 ff.

¹⁵⁰Vgl. Vukovic und Jakovljevic (2016), S. 314.

¹⁵¹Vgl. Janicijevic et al. (2013), S. 161.

¹⁵²Vgl. Mangione et al. (2001), S. 1051.

¹⁵³Vgl. Osiya et al. (2017), 1 f.

¹⁵⁴Vgl. Marshall und Hays (1994), S. 2.

¹⁵⁵Vgl. Thayaparan und Mahdi (2013), S. 1.

¹⁵⁶Vgl. Marshall und Hays (1994), S. 2.

Tabelle 6: Patientenzufriedenheitsdimensionen nach Vukovic et al.; Quelle: In Anlehnung an Vukovic et al. (2012), S. 718.

Dimension	Inhaltliche Definition
Sympathie	Sympathieniveau der Ärztin / des Arztes
Fachwissen	Empfundene Kompetenz der Ärztin / des Arztes
Behandlungszeit	Ausreichende Behandlungszeit
Information	Ausreichende Information
Wartezeit	Angemessene Wartezeit
Erreichbarkeit	Erreichbarkeit der Ärztin / des Arztes in Notfällen
Respekt	Rücksichtnahme auf die Patientenmeinung
Ernsthaftigkeit	Grad der Problemlösungskompetenz
Zwischenmenschliche Beziehung	Persönliche Eigenschaften der Ärztin / des Arztes
Zufriedenheit mit dem Praxispersonal	Generelle Zufriedenheit mit dem Praxispersonal
Gesamtzufriedenheit	Gesamtzufriedenheit mit den untersuchten Merkmalen

auf die behandelnde Ärztin / den behandelnden Arzt. Finanzielle Aspekte summieren die empfundene finanzielle Gerechtigkeit medizinischer Versorgung und die generelle Zufriedenheit beurteilt die Gesamtzufriedenheit mit der erhaltenen Behandlung.¹⁵⁷ Zusammenfassend stellt Tabelle 7 die unterschiedlichen Zufriedenheitsfaktorkategorien gegenüber.

Der PSQ 18 hat sich zu einem anerkannten Instrument im Bereich der Patientenzufriedenheitsmessung entwickelt, welches durch seine kompakte Beschaffenheit reproduzierbare Ergebnisse liefert und Erhebungen mit relativ geringem monetären und zeitlichem Aufwand ermöglicht.¹⁵⁸

4.2. Deutschsprachige Skalen zur Messung der Patientenzufriedenheit

Da davon ausgegangen werden muss, dass Patientenzufriedenheitsskalen, welche für den deutschsprachigen Markt entwickelt wurden, eher den jeweiligen regionalen oder politischen Bedingungen entsprechen, durch die Such- und Auswahlmethode jedoch internationale Messskalen bzw. solche in englischer Sprache begünstigt wurden, sollen hier zusätzlich zwei deutschsprachige Skalen zur Erhebung der Patientenzufriedenheit mit Hausärztinnen und -ärzten vorgestellt werden. Diese sind die Zufriedenheit in der Arztpraxis aus Patientenperspektive Skala (ZAP) sowie die Qualitätspraxen Skala (QP). Tabelle 8 fasst die beiden Skalen hinsichtlich Fragen bzw. Item-Anzahl, Antwortskalenformat, Veröffentlichungsjahr, Entwicklungs- und Erstverwendungsort und Anzahl an gemessenen Zufriedenheitsfaktorkategorien bzw. Dimensionen zusammen.

¹⁵⁷Vgl. Osiya et al. (2017), 1 f.

¹⁵⁸Vgl. Thayaparan und Mahdi (2013), S. 1.

4.2.1. Zufriedenheit in der Arztpraxis aus Patientenperspektive Skala (ZAP)

Die Zufriedenheit in der Arztpraxis aus Patientenperspektive Skala (ZAP) wurde in einem mehrstufigen Verfahren entwickelt, um die prozessbezogene Patientenzufriedenheit im Rahmen des Qualitätsmanagements von Hausarztpraxen zu messen. Der Fragebogen besteht aus insgesamt 23 Punkten, welche auf vier und fünfstufigen Likert-Skalen zu beantworten sind. Zur Konstruktion des Erhebungsinstrumentes wurde zunächst ein interimistischer Fragebogen entwickelt, welcher relativ viele Zufriedenheitsdimensionen berücksichtigte und in zwei Patientenbefragungsdurchläufen getestet wurde. Mit Hilfe statistischer Methoden wurden daraufhin wenig einflussnehmende Zufriedenheitsdimensionen eliminiert und das optimierte Befragungsinstrument editiert.¹⁵⁹

Insgesamt werden mit der ZAP fünf verschiedene Faktorkategorien der Patientenzufriedenheit gemessen. Diese sind die Arzt-Patienten-Interaktion, die Information der Patientinnen und Patienten, die Wirksamkeit der Behandlung, die empfundene fachliche Kompetenz der behandelnden Hausärztin / des behandelnden Hausarztes und die Praxisorganisation.¹⁶⁰ Wie Tabelle 9 zusammenfasst, wird unter der Arzt-Patienten- Interaktion das Verständnis und Einfühlungsvermögen und unter Information die Informationsvermittlungskompetenz der behandelnden Hausärztin / des behandelnden Hausarztes verstanden. Die fachliche Kompetenz summiert Faktoren, welche die medizinische Qualifikation bewerten und die Praxisorganisation beschäftigt sich bspw. mit der Wartezeit in der Praxis und der Freundlichkeit des Praxispersonals. Abschließend misst die Behandlungswirksamkeit die Gesamtzufriedenheit mit der Behandlung und Genesung.¹⁶¹

¹⁵⁹Vgl. Bitzer et al. (1999), S. 196 ff.

¹⁶⁰Vgl. Berger und Lenz (2006), S. 18.

¹⁶¹Vgl. Bitzer et al. (1999), S. 202 f.

Tabelle 7: Patientenzufriedenheitsdimensionen nach Osiya et al.; Quelle: In Anlehnung an Osiya et al. (2017), 1 f.

Dimension	Inhaltliche Definition
Kommunikation	Vermittlungs- und Verständnissfähigkeit
Zwischenmenschliche Belange	Persönlicher Umgang
Erreichbarkeit und Komfort	Zugang, Wartezeit und Wahlmöglichkeit
Technische Qualität	Technische Ausstattung und fachliche Kompetenz
Konsultationsdauer	Nachhaltige und stressfreie Untersuchungssituation
Finanzielle Aspekte	Behandlungs- und Folgekosten
Generelle Zufriedenheit	Gesamtzufriedenheit mit medizinischer Erstversorgung

Tabelle 8: Ausgewählte deutschsprachige Skalen zur Erhebung der Patientenzufriedenheit; Quelle: Verfasser.

Skala	Fragen	Skalenformat	Veröffentlicht	Entwickelt für	Dimensionen / Überkategorien
ZAP	23	4/5-stufig / Likert	1999	DE	5
QP	45	verschiedene	2004	DE	6

Tabelle 9: Patientenzufriedenheitsdimensionen nach Bitzer et al.; Quelle: In Anlehnung an Bitzer et al. (1999), S. 196 ff.

Faktorkategorie	Inhaltliche Definition
Arzt-Patienten-Interaktion	Persönlicher Umgang
Information	Ausreichende Information
Wirksamkeit der Behandlung	Gesamtzufriedenheit mit medizinischer Erstversorgung
Fachliche Kompetenz	Empfundene Kompetenz der Ärztin / des Arztes
Praxisorganisation	Merkmale der Praxisstruktur

4.2.2. Qualitätspraxen Skala (QP)

Die Qualitätspraxen Skala (QP) wurde entwickelt, um wahrgenommene Qualitätsdimensionen in Hausarztpraxen zu messen.¹⁶² Der Fragebogen wurde auf Grundlage eines eingehenden Literaturstudiums und des verbalen Austauschs mit teilnehmenden Allgemeinmedizinern, Ordinationsmitarbeitern sowie Mitgliedern der European Foundation of Quality Management entwickelt. Mit Hilfe der QP werden unterschiedliche Qualitäts- und Zufriedenheitsaspekte sowohl auf 4-stufigen Likert als auch mit dichotomen nominalen Antwortskalen gemessen.¹⁶³ Die Skala besteht aus 45 Fragepunkten, welche sich auf 6 verschiedene Qualitäts- bzw. Zufriedenheitsdimensionen beziehen, wobei 3 der Überkategorie Praxis und 3 der Kategorie Behandlung zugeordnet sind.¹⁶⁴ Zu den mit dem Befragungsinstrument untersuchten Qualitäts- und Zufriedenheitsfaktorkategorien zählen die Praxisausstattung, die Terminvereinbarung, die Praxisorganisation, Kompetenz und Information, Interaktion und Kommunikation sowie die emotionale und persönliche Un-

terstützung.¹⁶⁵ In der Praxisausstattungsdimension wird das Praxis-Interior bewertet und unter Terminvereinbarung ist der prozessuale Ablauf und die Zufriedenheit mit der Terminvergabe zu verstehen. Die Praxisorganisation bewertet die prozessualen Praxisabläufe und die Kompetenz- und Informationsdimension bewertet wiederum das Fachwissen und die Informationsbereitschaft der behandelnden Ärztin / des behandelnden Arztes. Die letzten beiden Dimensionen, Interaktion und Kommunikation sowie die emotionale und persönliche Unterstützung, beschäftigen sich mit dem ärztlichen Einfühlungsvermögen und zwischenmenschlichen Aspekten.¹⁶⁶ Die Definitionen der untersuchten Qualitäts- und Zufriedenheitsfaktorkategorien sind in Tabelle 10 übersichtlich zusammengefasst.

4.3. Resümee

Um die Untersuchungskategorien der einzelnen Patientenzufriedenheitsskalen vergleichbar zu gestalten, wurden

¹⁶²Vgl. Nübling et al. (2004), S. 301.

¹⁶³Vgl. Niebling et al. (2003), S. 531.

¹⁶⁴Vgl. Nübling et al. (2004), S. 301 f.

¹⁶⁵Vgl. Berger und Lenz (2006), S. 18.

¹⁶⁶Vgl. Niebling et al. (2003), S. 532 f.

Tabelle 10: Patientenzufriedenheitsdimensionen nach Nübling et al.; Quelle: In Anlehnung an Nübling et al. (2004), S. 301 ff.

Faktorkategorie	Inhaltliche Definition
Praxisausstattung	Räumliche Gegebenheiten
Terminvereinbarung	Prozess der Terminvergabe
Praxisorganisation	Prozessuale Praxisabläufe
Kompetenz und Information	Fachwissen und Informationsbereitschaft
Interaktion und Kommunikation	Vermittlungs- und Verständnisfähigkeit
Emotionale und persönliche Unterstützung	Zwischenmenschliche Aspekte

diese normiert und jeweils einer der Dienstleistungsdimensionen Potentialdimension, Prozessdimension oder Ergebnisdimension zugeordnet.¹⁶⁷

Die Potentialdimension beinhaltet hierbei jene Faktoren, die bereits vor Behandlungsbeginn bzw. Praxisbesuch vorhanden sein müssen, um innerhalb des Behandlungsprozesses Zufriedenheit auszulösen und summieren damit die Beschaffenheit der objektiven und subjektiven Leistungsvoraussetzungen.¹⁶⁸ Im Gegensatz dazu beschreibt die Prozessdimension jene Faktoren, die innerhalb des Leistungserstellungsprozesses aktiviert werden. Diese entstehen in der Regel durch den Einsatz der Potentialfaktoren, also durch die Inanspruchnahme der Vorleistung und können überwiegend nicht vollkommen beeinflusst werden, da die Kundin / der Kunde oder die Patientin / der Patient stets eigene Einflussfaktoren in den Leistungserstellungsprozess miteinbringt.¹⁶⁹ Abschließend wird innerhalb der Ergebnisdimension das Gesamtergebnis bzw. die Gesamtzufriedenheit mit dem Ergebniserstellungsprozess oder die Evaluierung der als angemessen oder unangemessen empfundenen Gegenleistung beurteilt.¹⁷⁰

Tabelle 11 stellt die unterschiedlichen internationalen Patientenzufriedenheitsskalen, gruppiert nach Potentialdimension, Prozessdimension und Ergebnisdimension, gegenüber. Die diesbezügliche Kategorisierung und Gruppierung der einzelnen Zufriedenheitsdimensionen wurde durch folgendes Prozedere sichergestellt: Ware et al. beschreiben bspw. die technische Qualität als fachliche Anbieterkompetenz.¹⁷¹ Da diese Definition sehr ähnlich jener der Zufriedenheitsdimensionen Medizinisch-technische Betreuung,¹⁷² Technische Aspekte,¹⁷³ und Fachwissen¹⁷⁴ ist, wurden diese Dimensionen innerhalb der Kategorie Fachliche Kompetenz zusammengelegt. Eine ähnliche Verfahrensweise wurde unter anderem auch bzgl. der Dimensionen Arzt-Patienten-

Beziehung sowie Auskunft und Hilfestellung,¹⁷⁵ Vertrauen und Patientenkenntnis¹⁷⁶ oder Sympathie angewandt.¹⁷⁷

Die unterschiedlichen Zufriedenheitsdimensionen wurden innerhalb der Kategorie Zwischenmenschliche Faktoren zusammengefasst. Diese zusammenfassende Normierungsmethodik wurde nachfolgend bei der Untersuchung aller Skalen und Dimensionen durchgeführt. Zur Kategorisierung der Zufriedenheitsdimensionen wurde eine eigene Auswertungsspalte erstellt, welche die jeweiligen Faktoren oder Faktorenbündel entweder der Potentialdimension, der Prozessdimension oder der Ergebnisdimension zuordnet. Die diesbezüglichen Zuordnungstabellen sind im Anhang ab Seite 104 zu finden.

In Tabelle 11 ist unter anderem zu erkennen, dass im Bereich der Potentialdimension die fachliche Kompetenz der behandelnden Ärztin / des behandelnden Arztes und die Praxisorganisation von allen internationalen Skalen gemessen wird. Innerhalb der Prozessdimension sind hingegen die zwischenmenschlichen Aspekte, die Behandlungsdauer und Wartezeit sehr stark vertreten. Eine Sonderstellung nimmt hier die verbale Verständigung ein, da diese von allen Skalen gemessen wird, jedoch nur bei dreien explizit als eigenständiger Zufriedenheitsfaktor genannt und bei den anderen beiden als Teilbereich der zwischenmenschlichen Aspekte gemessen wird. Die Zufriedenheit mit dem Behandlungsergebnis, bzw. die generelle Zufriedenheit in Bezug auf die untersuchten Merkmale, wird von vier Skalen ermittelt und die Zufriedenheit mit der finanziellen Gegenleistung für die erhaltene Behandlung wird als Aspekt der Ergebnisdimension nur innerhalb der beiden amerikanischen Skalen PSQF2 und PSQ 18 gemessen.

Tabelle 12 stellt nun die beiden untersuchten deutschsprachigen Skalen gegenüber, wobei die Normierung und Kategorisierung der einzelnen Zufriedenheitsdimensionen und Aspekte nach derselben Verfahrensweise erfolgte wie jene der internationalen Skalen. Die diesbezüglichen Zuordnungstabellen sind im Anhang ab Seite 104 zu finden.

Im Bereich der deutschsprachigen Skalen herrscht innerhalb der Potentialdimension Übereinstimmung bei der Unter-

¹⁶⁷Diesbezügliche Detailinformationen sind im Anhang ab Seite 104 zu finden.

¹⁶⁸Vgl. Meffert und Bruhn (2012), S. 89.

¹⁶⁹Vgl. Benkenstein (2008), S. 254 f.

¹⁷⁰Vgl. Bruhn (2008), S. 50.

¹⁷¹Vgl. Ware et al. (1983), S. 247.

¹⁷²Vgl. Grol et al. (1999), S. 5.

¹⁷³Vgl. Campbell et al. (2001), S. 91.

¹⁷⁴Vgl. Vukovic et al. (2012), S. 712.

¹⁷⁵Vgl. Grol et al. (1999), S. 5.

¹⁷⁶Vgl. Campbell et al. (2001), S. 91.

¹⁷⁷Vgl. Vukovic et al. (2012), S. 712.

Tabelle 11: Zufriedenheitsdimensionen der untersuchten internationalen Skalen; Quelle: Verfasser.

	PSQF2	EUROPEP	GPAS	QPS	PSQ 18
Potentialdimension					
Fachliche Kompetenz	✓	✓	✓	✓	✓
Organisation	✓	✓	✓	✓	✓
Ausstattung	✓	✓			✓
Umgebungsmerkmale	✓				
Kontinuität	✓		✓		
Prozessdimension					
Zwischenmenschliche Aspekte	✓	✓	✓	✓	✓
Behandlungsdauer / Wartezeit	✓	✓	✓	✓	✓
Verbale Verständigung			✓	✓	✓
Praxispersonal			✓	✓	
Ergebnisdimension					
Behandlungsergebnis	✓		✓	✓	✓
Finanzielle Faktoren	✓				✓

Tabelle 12: Zufriedenheitsdimensionen der untersuchten deutschsprachigen Skalen; Quelle: Verfasser.

	ZAP	QP
Potentialdimension		
Fachliche Kompetenz	✓	✓
Praxisorganisation	✓	✓
Praxisausstattung		✓
Terminvergabe		✓
Prozessdimension		
Arzt-Patient-Interaktion und Kommunikation	✓	✓
Information	✓	✓
Ergebnisdimension		
Wirksamkeit der Behandlung	✓	

suchung der empfundenen Zufriedenheit mit der fachlichen Kompetenz der behandelnden Ärztin / des behandelnden Arztes und der Praxisorganisation. Auch innerhalb der Prozessdimension, welche in diesem Fall die Zufriedenheit mit der Arzt- Patient-Interaktion und Kommunikation sowie der Informationsbereitschaft der behandelnden Ärztin / des behandelnden Arztes abbildet, werden alle erwähnten Aspekte von den analysierten Skalen gemessen. Unterschiede gibt es, wie schon bei den betrachteten internationalen Skalen, bei der Ergebnisdimension, bei welcher die Zufriedenheit mit der Wirksamkeit der Behandlung nur von einer Skala gemessen wird.

Wie die Gegenüberstellung der einzelnen Skalen zeigt, sind diese durch Einigkeit, Abweichungen und Diskrepanzen bei der Definition und Messung unterschiedlicher Zufriedenheitsfaktoren gekennzeichnet, wobei dieser Umstand natürlich auch auf unterschiedliche Untersuchungsschwerpunkte

zurückzuführen ist. Zu den am häufigsten untersuchten Zufriedenheitsfaktoren zählen die ärztliche Fachkompetenz, die Praxisorganisation und zwischenmenschliche Aspekte. Untersuchungsschwächen bestehen möglicherweise noch im Bereich der Ergebnisdimension.

5. Methoden zur Messung von Patientenzufriedenheit

Die Bestimmung des Niveaus der Patientenzufriedenheit ermöglicht Ärztinnen und Ärzten ihre persönlichen Merkmale, aber auch Merkmale wie die Praxisorganisation oder beeinflussbare äußere Umstände, den Patientenwünschen entsprechend anzupassen. Derartige Adaptionen haben nicht nur eine positive Wirkung auf die Wettbewerbsposition der Ärztin / des Arztes und der Praxis, sondern können auch

positive Behandlungseffekte erzielen.¹⁷⁸ Vor diesem Hintergrund erscheint es erstrebenswert, Möglichkeiten zur Evaluation der subjektiv empfundenen Patientenzufriedenheit abzubilden.¹⁷⁹

Die Basis jeder marktrelevanten Untersuchung bilden Informationen. Um Informationen über ein bestimmtes Untersuchungsgebiet zu erhalten, können diese entweder selbst erhoben (Primärdaten) oder interne und externe Quellen (Sekundärdaten) nach dem Aufkommen adäquater Daten durchsucht werden.¹⁸⁰ Sekundärdaten bieten die Möglichkeit, Erkenntnisse aus bereits bestehenden Datenquellen zu gewinnen und sind dadurch gekennzeichnet, dass sie bereits in der Vergangenheit durch interne oder externe Untersuchungen erhoben wurden. Grundsätzlich sollte jede Art von Datensammlung mit der Suche nach Sekundärdaten beginnen, denn diese sind in der Regel kostengünstiger und einfacher zu beschaffen als Primärdaten und bestimmte Informationen können über eine Primärdatenerhebung nicht oder nur mit sehr hohem Ressourcenaufwand beschafft werden. Darüber hinaus unterstützen Sekundärdaten die Planung und Entwicklung von Primärdatenerhebungen bzw. Erhebungsinstrumenten.¹⁸¹ Der Umgang mit Sekundärdaten ist demnach dadurch gekennzeichnet, dass diese, statt selbst erhoben, von bestehenden Datensammlungen abgeleitet werden. Jene Datensammlungen können sowohl unternehmensintern, bspw. durch die Verwendung von Daten aus früheren Untersuchungen, als auch unternehmensextern zum Vorschein kommen.¹⁸²

Die Konzeption von Sekundärdatenanalysen setzt präzise und messbare Problem- und Fragestellungen voraus, welche möglichst exakt und detailliert skizziert sein sollen. Diese Vorgehensweise gilt als grundlegende Anforderung, um das Studiendesign und die extrahierten Daten bewerten und Zeit und Kosten kalkulieren zu können. Somit werden erst durch die Festlegung der Fragestellung der eigenen Untersuchung Komponenten der Sekundärdatenanalyse, wie bspw. die zu untersuchende Gruppe oder die Abgrenzung spezieller Untersuchungsvariablen, bestimmt.¹⁸³ Um der eigenen Problem- und Fragestellung gerecht zu werden und das angestrebte Forschungsziel zu erreichen, ist es unabdingbar, nach Möglichkeit sämtliche Informationsquellen zu ermitteln und nach adäquaten Informationen zu durchsuchen. Diese Vorgehensweise ist die Grundlage, um die gesammelten Daten im nächsten Schritt zu sichten, zu bewerten und zu archivieren.¹⁸⁴

Die Sekundärmarktforschung beinhaltet unterschiedliche Vorteile sowie relativ geringe Kosten,¹⁸⁵ hat aber, je nach Fragestellung, auch Grenzen und kann unter Umständen mit Problemen verbunden sein. So kann es vorkommen, dass

für die spezifische Fragestellung keine Sekundärdaten vorhanden oder auffindbar sind, die verwendeten Skalen nicht mit der eigenen Untersuchung kompatibel sind, die Daten nicht aktuell sind, oder Objektivität, Reliabilität und Validität nicht in suffizientem Ausmaß berücksichtigt wurden. Sekundärdaten bieten eine Vielzahl unterschiedlicher Anwendungsbereiche. Einerseits kann die Sekundärdatenerhebung die Primärdatenerhebung gänzlich ersetzen. Diese Anwendungsform kommt jedoch nur zustande, wenn die gesammelten Daten das Erkenntnisinteresse vollends stillen. Andererseits werden die Daten häufig genutzt, um die Primärdatenerhebung vorzubereiten. Hierbei entwickelt sich die Primärdatenerhebung aus den Daten der Sekundärdatenanalyse. Die dritte häufige Anwendungsform besteht in der Ergänzung der Primärdaten durch gesammelte Sekundärdaten. Anwendung findet diese Vorgehensweise bspw. indem Primärdaten mit Sekundärdaten verglichen werden.¹⁸⁶

Die Verwendung externer oder interner Datenquellen ist grundsätzlich durch einen geringeren Kostenaufwand gekennzeichnet als die Neuerhebung dieser Daten.¹⁸⁷

Insbesondere die externen Datenquellen können hierbei, je nach Fragestellung, sehr vielfältig sein und reichen von Fachzeitschriften und Büchern, über Internetquellen, wie Foren oder Datenbanken, bis hin zu Quellen von Marktforschungsinstituten, öffentlichen Gremien oder Wettbewerbern, welche dieselbe oder eine ähnliche Fragestellung bereits untersucht haben.¹⁸⁸

Im Gegensatz dazu wären interne Datenquellen im Bereich der internen Datenbanken sowie internen Statistiken, Berichten oder Beschwerdesystemen zu finden.¹⁸⁹ Interne und externe Daten können, je nach Zielsetzung, aus einer Vielzahl unterschiedlicher Quellen stammen. Aus diesem Grund gibt Tabelle 13 beispielhaft einen Überblick über mögliche interne und externe Datenquellen, welche der Messung von Patientenzufriedenheit dienlich sein können.¹⁹⁰

Bedingt durch ihre Beschaffenheit reichen Sekundärdaten oftmals nicht aus, um die Fragestellung effizient zu beantworten. Um mehr über die Patientenmerkmale in Erfahrung zu bringen empfiehlt es sich daher, Daten vor Ort zu erheben.¹⁹¹ Bei der Erhebung von Primärdaten kann generell zwischen Instrumenten der Befragung sowie Instrumenten der Beobachtung unterschieden werden, wobei Experimente Mischformen der beiden genannten Erhebungsformen darstellen.¹⁹² Eine weitere Unterteilung kann in dieser Hinsicht zwischen qualitativen und quantitativen Erhebungsmethoden getroffen werden.¹⁹³

Qualitative Erhebungsmethoden versuchen einen Themenbereich in seiner Gesamtheit zu betrachten und führen

¹⁷⁸Vgl. Gericke et al. (2004), S. 730.

¹⁷⁹Vgl. Bieber et al. (2011) S. 80.

¹⁸⁰Vgl. Homburg und Krohmer (2006), S. 262.

¹⁸¹Vgl. Berekoven et al. (2007), S. 40.

¹⁸²Vgl. Homburg und Krohmer (2006), S. 262.

¹⁸³Vgl. Swart et al. (2008), S. 54 ff.

¹⁸⁴Vgl. Magerhans (2016), S. 64 ff.

¹⁸⁵Vgl. Olbrich et al. (2012), S. 67.

¹⁸⁶Vgl. Magerhans (2016), S. 64 ff.

¹⁸⁷Vgl. Olbrich et al. (2012), S. 68.

¹⁸⁸Vgl. Berekoven et al. (2007), S. 39 ff.

¹⁸⁹Vgl. Magerhans (2016), S. 69.

¹⁹⁰Vgl. Kuß (2012), S. 45.

¹⁹¹Vgl. Pirk und O. (2012), S. 201 f.

¹⁹²Vgl. Homburg und Krohmer (2006), S. 262.

¹⁹³Vgl. Brosius et al. (2012), S. 3.

Tabelle 13: Interne und externe Sekundärdatenquellen; Quelle: In Anlehnung an Magerhans (2016), S. 65 und die dort zitierte Literatur und Homburg und Krohmer (2006), S. 269.

Interne Sekundärdatenquellen	Externe Sekundärdatenquellen
Kunden- bzw. Patientenakte	Statistiken und amtliche Institutionen
Besuchsberichte	Veröffentlichungen von wissenschaftlichen Institutionen
Kunden- bzw. Patientendaten	Veröffentlichungen von Verlagen
CRM-Daten	Veröffentlichungen im Internet

in der Regel nicht zu quantifizierbaren Ergebnissen.¹⁹⁴ Das bedeutet, dass qualitative Methoden vor allem dann vorteilhaft eingesetzt werden können, wenn das Untersuchungsfeld noch relativ unbekannt ist.¹⁹⁵ Im Gegensatz dazu wird mit Hilfe von quantitativen Erhebungsmethoden versucht, faktische Relationen in zahlenmäßiger Struktur auf Basis einer numerischen Erhebung darzustellen. Ziel ist es demnach, die Realität so quantifizierbar wie möglich zu gestalten.¹⁹⁶

Eine wichtige Einflussgröße in Bezug auf die Messung von Patientenzufriedenheit ist der Zeitpunkt der Befragung. Dieser kann grundsätzlich zu Beginn der Behandlung oder nach der Behandlung stattfinden und führt naturgemäß zu divergenten Untersuchungsergebnissen.¹⁹⁷ Die Befragung vor Behandlungsbeginn eignet sich gut, um Diskrepanzen in der Wahrnehmung unterschiedlicher Patientinnen und Patienten zu ermitteln und diese vor der Auswertung zu kompensieren, ist aber in Bezug auf die Messung der Zufriedenheit mit der Ärztin / dem Arzt und der Behandlung durch Beurteilungsdefizite gekennzeichnet.¹⁹⁸ Werden hingegen Patientinnen und Patienten am Ende der Behandlung befragt, so können diese ihre jeweilige aktuelle Zufriedenheit zwar durchaus beurteilen, hierbei wird jedoch oftmals die individuelle Wahrnehmung zu hoch bzw. zu minder bemessen, was bedeutet, dass die Befragung zu einem späteren Zeitpunkt zu divergenten Ergebnissen führen könnte.¹⁹⁹ Aus diesem Grund wird vermehrt davon ausgegangen, dass sich der optimale Patientenbefragungszeitpunkt in Bezug auf die Zufriedenheit mit den Behandlungsumständen und den behandelnden Ärztinnen und Ärzten zwischen einer und vier Wochen nach der Behandlung bewegt, da der emotionale Einfluss auf die Wahrnehmung bis dahin abgeschwächt sein sollte, die Behandlung aber noch nicht so lange her ist, dass Erinnerungsdefizite auftreten.²⁰⁰ Neben der Befragung aktueller Patientinnen und Patienten kann es auch von Interesse sein, potentielle oder ehemalige Patientinnen und Patienten in die Untersuchung miteinzubeziehen.²⁰¹ Diese Vorgehensweise erlaubt es den jeweiligen Ärztinnen und Ärzten, Anpassungen in Bezug

auf die Bedürfnisse der befragten Patientinnen und Patienten durchzuführen und dadurch potentielle Patientinnen und Patienten zu aktuellen zu machen, aktuelle Patientinnen und Patienten zu binden und ehemalige wiederzugewinnen.²⁰² Die Überlegungen hinsichtlich des optimalen Patientenbefragungszeitpunktes sind in Abbildung 6 übersichtlich zusammengefasst.

Da Vollerhebungen, also die Befragung aller Patientinnen und Patienten einer Arztpraxis, in den meisten Fällen nicht wirtschaftlich durchführbar sind, werden in der Regel Stichproben herangezogen, um das jeweilige Erkenntnisinteresse zu stillen, wobei hier wiederum zwischen zufälligen und nicht zufälligen Auswahlverfahren unterschieden wird.²⁰³ Zufällige Auswahlverfahren summieren unterschiedliche Ansätze, bei denen jedes Untersuchungsobjekt, je nach gewähltem Verfahren, die mehr oder weniger gleiche Wahrscheinlichkeit besitzt, in die Stichprobe miteinbezogen zu werden. Hierbei gilt, dass die Werte der Stichprobe sich stärker den zu erwartenden Werten in der Grundgesamtheit annähern je größer die Stichprobe ist.²⁰⁴ Bei nicht zufälligen Auswahlmethoden werden hingegen die Untersuchungsobjekte nach bestimmten Kriterien ausgewählt, wodurch jedoch häufig Problemfelder hinsichtlich der Repräsentativität der Forschungsergebnisse entstehen, da die Selektion nicht nach der Maxime der Wahrscheinlichkeit, sondern nach vordefinierten Überlegungen erfolgt.²⁰⁵

Ist die Grundgesamtheit, bspw. Patientinnen und Patienten einer Arztpraxis, definiert, wird entschieden, wie groß die Stichprobe sein soll bzw. welche Elemente aus der Grundgesamtheit befragt werden sollen.²⁰⁶ Insbesondere im Bereich der Patientenzufriedenheitsmessung ist darauf zu achten, dass die Stichprobe als reine Zufallsstichprobe durchgeführt wird oder die ausgewählten Untersuchungsobjekte auch ihrem jeweiligen Vorkommen in der Grundgesamtheit entsprechen, da das empfundene Zufriedenheitsniveau auch in Zusammenhang mit soziodemografischen Aspekten steht.²⁰⁷ Zu den einflussnehmenden soziodemografischen Ausprägungen zählen in diesem Zusammenhang unter anderem das Geschlecht, das Alter, der Bildungsstand, die familiäre

¹⁹⁴Vgl. Häder (2015), S. 15.

¹⁹⁵Vgl. Brüsemeister (2008), S. 19.

¹⁹⁶Vgl. Schwaiger und Zimmermann (2011), S. 420 f.

¹⁹⁷Vgl. Keiner (2006), S. 99.

¹⁹⁸Vgl. Bortz und Döring (2016), S. 102 f.

¹⁹⁹Vgl. Frodl (2011), S. 62.

²⁰⁰Vgl. Neugebauer und Porst (2001), S. 19.

²⁰¹Vgl. Töpfer (2008b), S. 386.

²⁰²Vgl. Gericke et al. (2004), S. 730.

²⁰³Vgl. Löffler und Von der Heyde (2014), S. 19.

²⁰⁴Vgl. Berekoven et al. (2007), S. 46.

²⁰⁵Vgl. Brosius et al. (2012), S. 69.

²⁰⁶Vgl. Weis und Steinmetz (2008), S. 88.

²⁰⁷Vgl. Fenton et al. (2012), S. 403.

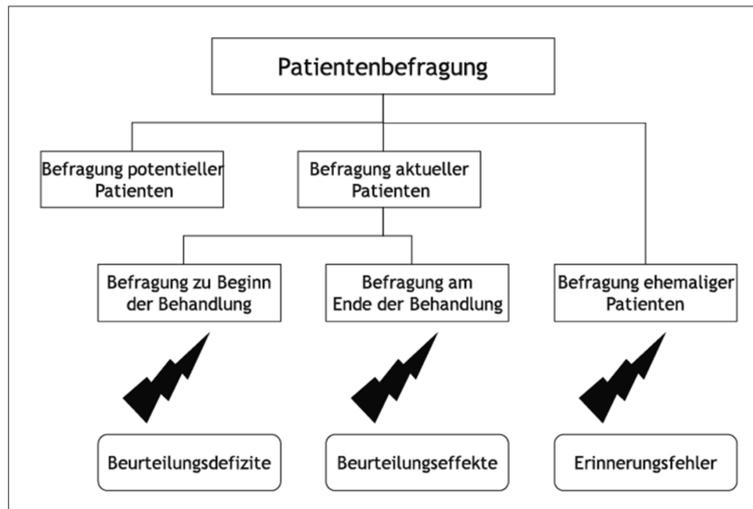


Abbildung 6: Übersicht Patientenbefragungszeitpunkte; Quelle: Frodl (2011), S. 62.

Situation sowie die Nationalität der Patientinnen und Patienten.²⁰⁸ Die Merkmalsausprägungen einzelner Patientinnen und Patienten sind nicht zu verkennen, da Untersuchungen, wie beispielsweise die zusammenfassende Metaanalyse von Salisbury et al., gezeigt haben, dass ein systematischer Zusammenhang zwischen Patienteneigenschaften wie Alter oder Ethnie und empfundener Zufriedenheit mit Ärztinnen und Ärzten besteht, dieser Zusammenhang jedoch in der Anpassung der Behandlungsmethoden häufig vernachlässigt wird.²⁰⁹

Neben soziodemografischen Merkmalen kann auch das Krankheitsbild selbst die Zufriedenheit beeinflussen. Einerseits fällt es in der Regel schwerer, Patientinnen und Patienten mit insuffizientem Gesundheitszustand zu erreichen und es stellt sich je nach Erkrankungsgrad die Frage, ob diese geistig oder körperlich fähig sind, die Fragen zu beantworten. Andererseits erzielt physische Gesundheit einen positiven Effekt in Bezug auf die Zufriedenheitsbeurteilung. Demnach ist davon auszugehen, dass Patientinnen und Patienten mit einer schweren Erkrankung ihre Zufriedenheit grundsätzlich anders beurteilen als jene mit einer leichten Erkrankung.²¹⁰

Die Auswahl der Untersuchungsmethode bzw. des Untersuchungsinstrumentes hängt insbesondere im medizinischen Bereich von unterschiedlichen Faktoren in Bezug auf die Patientenfreundlichkeit ab. So muss speziell bei Zufriedenheitsbefragungen im Gesundheitsumfeld darauf geachtet werden, dass das Instrument geeignet ist, um das relativ komplexe Untersuchungsumfeld verständlich darzustellen. Davon abgesehen soll die Untersuchung im Einzelfall dennoch mit möglichst geringem Zeitaufwand verbunden sein.²¹¹

Die Analyse von persönlichen Merkmalen, Eigenschaften, Ansprüchen, Forderungen und subjektiver Zufriedenheit

ist generell ausschließlich über ein Befragungsinstrument durchführbar. Dementsprechend sind auch die Erhebung der aktuellen Patientenzufriedenheit sowie die Evaluation unterschiedlicher Dimensionen nur unter Zuhilfenahme einer direkten oder indirekten Patientenbefragung möglich.²¹²

Patientenzufriedenheitsmessungen haben in der Vergangenheit kontinuierlich zugenommen und stellen ein relativ komplexes Konstrukt eines Marktforschungsprozesses dar. Allen Methoden und Instrumenten gemein ist der Anspruch auf Objektivität, Reliabilität und Validität, da nur Methoden und Instrumente, welche jene Gütekriterien miteinbeziehen, systematisch korrekte Erhebungen begünstigen.²¹³ Objektivität beschreibt in diesem Zusammenhang einen Untersuchungsprozess, bei welchem die / der Untersuchende keinen Einfluss auf das Messergebnis ausübt, die Untersuchung also auch bei Durchführung durch eine andere Person zum selben Ergebnis führt.²¹⁴ Unter Reliabilität wird die Zuverlässigkeit des Messergebnisses verstanden. Das bedeutet, dass die Untersuchung bei erneuter Durchführung zu vergleichbaren Ergebnissen führen soll, also frei von Messfehlern ist.²¹⁵ Validität hingegen liegt dann vor, wenn ein Messverfahren genau das misst, was gemessen werden soll.²¹⁶

In den folgenden Unterkapiteln werden unterschiedliche Datenerhebungsmethoden für die Untersuchung der Patientenzufriedenheit mit Ärztinnen und Ärzten dargestellt. Tabelle 14 gibt Auskunft über die vorgestellten Untersuchungen und Studien im Bereich der Erhebung von Primärdaten der Patientenzufriedenheit und fasst die wichtigsten Eckdaten der jeweiligen Erhebung zusammen.

²¹²Vgl. Neugebauer und Porst (2001), S. 19.

²¹³Vgl. Wüthrich-Schneider (2000b), S. 1119.

²¹⁴Vgl. Homburg und Krohmer (2006), S. 255.

²¹⁵Vgl. Moosbrugger und Kelava (2012), S. 11.

²¹⁶Vgl. Rammstedt (2010), S. 250.

²⁰⁸Vgl. Alrubaiee und Alkaa'ida (2011), S. 126.

²⁰⁹Vgl. Salisbury et al. (2010), S. 1 ff.

²¹⁰Vgl. Applebaum et al. (2004), S. 30.

²¹¹Vgl. Wüthrich-Schneider (2000b), S. 1118.

Tabelle 14: Ausgesuchte Patientenzufriedenheitserhebungen; Quelle: Verfasser.

Erhebungsmethode	Autoren	Teilnehmer	EI*	Jahr
Tiefeninterview	Joo et al.	13	N/A***	2017
Gruppendiskussion	Baker	> 40	CSQ	1990
C. I. T.**	Marcinowicz et al.	30	N/A***	2014
Vorschlagssystem	Mahler et al.	370	N/A***	2012
Mündliche Befragung	Stewart et al.	319	MOSSF36	2000
Schriftliche Befragung	Williams / Calnan	454	N/A***	1991
Telefonische Befragung	Stewart et al.	319	MOSSF36	2000
Onlinebefragung	Campbell et al.	> 2 Mio.	ENGPPS	2009
Beobachtung	McKinstry et al.	106	N/A***	2010
Experiment	Lomgo et al.	747	N/A***	2006
Onlineressourcen	Kadry et al	4999	N/A***	2011

* EI: Erhebungsinstrument
** C. I. T.: Critical Incident Technique. Vgl. [Matzler und Bailom \(2004\)](#), S. 267.
***N/A: Nicht angegeben (not available) bzw. nicht anwendbar. Vgl. [Bielińska \(2010\)](#), S. 227.

5.1. Befragung

Die Befragung ist die wohl am häufigsten angewandte Methode im Bereich der Patientenzufriedenheitsmessung.²¹⁷ Hierbei lassen sich eher qualitativ und eher quantitativ ausgerichtete Erhebungsmethoden unterscheiden.²¹⁸

Mit Hilfe von qualitativen Forschungsmethoden wird versucht, Erkenntnisse über Gegebenheiten in ihrer Gesamtheit zu erfassen.²¹⁹ Dabei liegt der Untersuchungsschwerpunkt nicht in der Quantifizierung des Analysegegenstandes, sondern der Fokus liegt darauf, facettenreiche und profunde Erkenntnisse zu erhalten.²²⁰ Qualitative Untersuchungsmethoden werden häufig angewandt, um Informationen zu erhalten, welche die Grundlage für anschließende repräsentative quantitative Erhebungen bilden. Dementsprechend definieren die diesbezüglich erhobenen Daten häufig den Untersuchungsrahmen und legen Aspekte und Komponenten des Untersuchungsgebietes fest.²²¹

Qualitative Untersuchungen werden demzufolge nicht angewandt, um genaue Messungen und Auswertungen durchzuführen, sondern dienen eher der Ideengewinnung und Hypothesenbildung. Dementsprechend sind die unterschiedlichen Methoden sehr vielfältig gestaltet, wobei sich indessen mehrere Übereinstimmungen erkennen lassen. So handelt es sich bei qualitativen Forschungsdesigns häufig um Untersuchungen mit marginaler gefestigter Struktur, innerhalb des Forschungsprozesses sind die Erhebung, die Analyse und die Hypothesenbildung eng verbunden, es handelt sich oftmals um nicht repräsentative Untersuchungen mit geringem Stichprobenumfang, die erhobenen Daten lassen sich

nicht oder nur schwer quantifizieren und es handelt sich stets um eine nicht oder kaum standardisierte Untersuchungsmethodik.²²²

Um ein Forschungsobjekt qualitativ zu untersuchen, reichen aufgrund der breiten Zielsetzung in der Regel relativ geringe Teilnehmerzahlen aus, wobei sich dieser Umstand erwartungsgemäß negativ auf die Repräsentativität der Untersuchung auswirken kann. Dieser Sachverhalt relativiert sich jedoch durch den prozessualen Charakter qualitativer Methoden, eher einen umfassenden Überblick und eine grobe Richtung vorzugeben als zu exakten Ergebnissen zu führen.²²³

Obwohl die Prämisse qualitativer Forschungsmethoden tendenziell auskunftsschaffend ausgerichtet ist, wird dennoch immer stärker darauf geachtet, dass die Untersuchungsergebnisse den Anspruch auf Validität und Reliabilität erfüllen. Dieser Anspruch mündet in vergleichbaren, wiederholbaren Untersuchungen und fundiert die logische Herleitung der Ergebnisse.²²⁴ Qualitative Befragungsmethoden bieten mittlerweile eine große Vielfalt an Instrumenten und Untersuchungsansätzen. Dazu zählen unter anderem die auch in dieser Arbeit vorgestellten Befragungsmethoden Tiefeninterview, Gruppendiskussion, Critical Incident Technique und die Analyse von Beschwerde- und Vorschlagssystemen.²²⁵

Im Gegensatz zu qualitativen Erhebungsmethoden sind quantitative Erhebungsmethoden dadurch gekennzeichnet, dass ausgewählte Attribute, auf Erfahrung beruhende Erkenntnisse, methodisch mit numerischen Werten belegt und quantitativ erfasst werden.²²⁶ Charakteristisch für den Umgang mit quantitativen Verfahren ist demnach die Erhe-

²¹⁷Vgl. [Wüthrich-Schneider \(2000b\)](#), S. 1118.

²¹⁸Vgl. [Homburg und Krohmer \(2006\)](#), S. 265.

²¹⁹Vgl. [Brosius et al. \(2012\)](#), S. 4.

²²⁰Vgl. [Kuß \(2012\)](#), S. 134.

²²¹Vgl. [Berekoven et al. \(2007\)](#), S. 89 f.

²²²Vgl. [Kuß \(2012\)](#), S. 136.

²²³Vgl. [Berekoven et al. \(2007\)](#), S. 89 f.

²²⁴Vgl. [Kuß \(2012\)](#), S. 136 f.

²²⁵Vgl. [Berekoven et al. \(2007\)](#), S. 92.

²²⁶Vgl. [Brosius et al. \(2012\)](#), S. 4.

bung von Informationen in Zahlenformat bzw. die Transformation von Faktoreigenschaften in zahlenmäßige Form. Die so gewonnenen Erkenntnisse werden generell unter Zuhilfenahme statistischer Auswertungsmethoden analysiert.²²⁷ Hierbei wird unter anderem untersucht, ob ein Untersuchungsgegenstand unterschiedliche Merkmalsausprägungen besitzt und wie sich diese zueinander verhalten, ob dieser Umstand auch auf andere Untersuchungsgegenstände zutrifft und wie sich jene Verhaltensweisen im Zeitverlauf verändern.²²⁸

Die Charakterisierung quantitativer Erhebungsmethoden beruht folglich darauf, dass die Erfassung und Analyse der gesammelten Informationen keiner Auslegung durch die / den Untersuchenden bedarf. Die gebündelten Daten sind allgemein plausibel und reproduzierbar. Dieser Umstand beschreibt zweifelsohne einen der erheblichsten Vorteile quantitativer Forschungsmethoden. Weitere positive Aspekte bestehen in der Regel darin, dass der Ablauf klar definiert und strukturiert ist und die Untersuchungsmethoden spezifisch charakterisierbar und abgrenzbar sind. Sowohl innerhalb von theoretischen als auch praktischen Problemfeldern kommen die Vorteile quantitativer Untersuchungsmethoden deutlich zum Vorschein, weshalb sie weitläufige Anwendungsbereiche innehaben und in unterschiedlicher Hinsicht eingesetzt werden.²²⁹

Quantitative Untersuchungsmethoden werden einerseits angewandt, um repräsentative Untersuchungsergebnisse zu erhalten, andererseits aber auch, um den durch qualitative Methoden erforschten Attributen, Merkmalen und Kriterien ihre jeweilige Stellung oder Positionierung zuzuordnen.²³⁰ Innerhalb der quantitativen Primärdatenerhebung wird zwischen normiertem mündlichem Interview, normierter schriftlicher Befragung, normiertem telefonischem Interview und der Onlinebefragung unterschieden.²³¹

Insbesondere bei schriftlichen Befragungsarten stellt die sogenannte Rücklaufquote eine wichtige Einflussgröße dar. Sie wird als Quotient des Rücklaufes, also Anteil der rückgesendeten Formulare im Gegensatz zu ausgesendeten Formularen, definiert. Traditionell werden häufig die Art und der Zeitpunkt der Befragung als determinierende Faktoren auf die Rücklaufquote definiert.²³² Einfluss auf die Rücklaufquote haben aber nicht nur Befragungsart und Zeitpunkt, sondern auch demografische Aspekte, das Interviewverhalten oder Anreize zur Beantwortung.²³³

Mazor et al. konnten darüber hinaus in ihrer Untersuchung aus dem Jahr 2002 belegen, dass die Rücklaufquote, insbesondere im Bereich der Patientenzufriedenheitsforschung, signifikant durch das Patientenzufriedenheitslevel beeinflusst wird. Innerhalb der Untersuchung war die Wahr-

scheinlichkeit, dass eine zufriedene Patientin / ein zufriedener Patient an der Untersuchung teilnimmt deutlich höher als die Wahrscheinlichkeit, dass eine unzufriedene / ein unzufriedener teilnimmt. Dieses Ergebnis deutet darauf hin, dass der Faktor Rücklaufquote Patientenzufriedenheitsmessungen erheblich beeinflussen kann, was unter Miteinbezug dieses Ergebnisses zu einer Überschätzung des Zufriedenheitsniveaus innerhalb der Untersuchungsgruppe führt. Diesem Schluss folgend könnte sogar davon ausgegangen werden, dass Ärztinnen und Ärzte mit einer schlechteren Leistung besser bewertet werden als jene mit einer besseren Leistung. Deutlich wird durch diese Darstellung jedenfalls, dass die Gegenüberstellung von Ärztinnen und Ärzten mit positiv wahrgenommener Performanz und solchen mit negativ wahrgenommener Performanz nicht aufgrund von unstrukturiert gesammelter Daten vollzogen werden kann.²³⁴ Bei der Primärdatenerhebung, insbesondere bei der Befragung, ist infolgedessen neben der bereits diskutierten Gütekriterien auch die Stichprobenqualität, demnach die Zusammensetzung der Stichprobe, welche sich aus der Rücklaufquote erheben lässt, zu beachten und zu beurteilen, um ein hohes Maß an Repräsentativität zu erlangen.²³⁵

Im Folgenden werden die eher qualitativen Untersuchungsmethoden (5.1.1 - 5.1.4), gefolgt von den eher quantitativen Befragungsmethoden (5.1.5) dargestellt. Insbesondere wird hierbei auf den Zusammenhang der jeweiligen Methode mit der Zufriedenheitsmessung mit Hausärztinnen und -ärzten eingegangen.

5.1.1. Tiefeninterview

Das Tiefeninterview bezeichnet ein verhältnismäßig ungebundenes qualitatives Erhebungsinstrument, welches als subjektives Gespräch Anwendung findet.²³⁶ Dabei handelt es sich um eine, einen längeren Zeitraum andauernde, tiefgehende Unterhaltung zwischen einer auskunftsfragenden und einer auskunftsgebenden Person über vorher festgelegte Themenbereiche, wobei die Interviewerin / der Interviewer das Gespräch derart zu führen versucht, dass sie / er nach Möglichkeit sämtliche für die auskunftsgebende Person bedeutsamen Charakteristika, Merkmale, Blickwinkel und Aspekte erfasst.²³⁷

Tiefeninterviews werden gewöhnlich durchgeführt, um grundlegende Erkenntnisse in Bezug auf die Wahrnehmung der Auskunftsperson zu erlangen,²³⁸ wobei diese Untersuchungsmethode auch dazu dient, den Befragten Beweggründe und Handlungsweisen bewusst zu machen, welche sie bislang nicht oder kaum erkannt haben, um eine korrespondierende Untersuchung umsetzen zu können.²³⁹ Abgesehen von den allgemeinen Auskünften der interviewten Personen werden auch Vorschläge in Bezug auf die Lösung identifizierter

²²⁷Vgl. Schwaiger und Zimmermann (2011), S. 421.

²²⁸Vgl. Brüsemeister (2008), S. 20.

²²⁹Vgl. Schwaiger und Zimmermann (2011), S. 421.

²³⁰Vgl. Berekoven et al. (2007), S. 89 f.

²³¹Vgl. Homburg und Krohmer (2006), S. 268.

²³²Vgl. Bischoff (2002), S. 8.

²³³Vgl. Wüthrich-Schneider (2000b), S. 1118.

²³⁴Vgl. Mazor et al. (2002), S. 1403

²³⁵Vgl. Tachtsoglou und König (2017), S. 268.

²³⁶Vgl. Homburg und Krohmer (2006), S. 264.

²³⁷Vgl. Salcher (1995), S. 34.

²³⁸Vgl. Homburg und Krohmer (2006), S. 264.

²³⁹Vgl. Kuß (2012), S. 132.

Problemfelder examiniert. Von substantieller Bedeutung ist in diesem Zusammenhang, dass der auskunftsgewährenden Person umfassende Gestaltungsmöglichkeiten in Bezug auf die Äußerung individueller Vorstellungen, Gefühle und Anliegen gewährt werden.²⁴⁰

Durch das offene Gespräch mit aktuellen, potentiellen und ehemaligen Patientinnen und Patienten dienen Tiefeninterviews den teilstrukturellen Zufriedenheitsmessungen und ermöglichen es Ärztinnen und Ärzten, einen ersten Überblick über Wünsche, Vorstellungen und Erwartungen ihrer Patientinnen und Patienten zu erhalten.²⁴¹ Die Methode eignet sich also insbesondere als Vorstudie für Untersuchungsbereiche, in welchen keine oder nur wenige Erkenntnisse vorliegen oder für Themenbereiche, welche unter alternativen Aspekten betrachtet werden sollen.²⁴²

Eine relativ aktuelle Studie in diesem Zusammenhang ist jene von Joo et al., bei welcher sowohl die Zufriedenheit von Patientinnen und Patienten mit dem Primary Care Projekt als auch einzelne Zufriedenheitsdimensionen mit Hausärztinnen und -ärzten untersucht und charakterisiert wurden. Die Studie wurde mit Hilfe qualitativer Daten, welche im Zuge von 13 halbstrukturierten Tiefeninterviews erhoben wurden, durchgeführt und konnte einen relativ starken Zusammenhang zwischen Patientenzufriedenheit und einer positiven Arzt-Patienten-Beziehung sowie der verbalen und nonverbalen Bereitstellung umfassender Informationen und Patientenzufriedenheit konstatieren.²⁴³

5.1.2. Gruppendiskussion

Unter einer Gruppendiskussion wird innerhalb der qualitativen Forschungsmethoden ein konzeptioneller Dialog verstanden, welcher unter der Prämisse geführt wird, Eigenschaften, Haltungen und Merkmale in Bezug auf ein bestimmtes, durch die Forschungsfrage festgelegtes, Untersuchungsareal zu bestimmen.²⁴⁴ Im Gegensatz zum Tiefeninterview, welches darauf abzielt, umfassende Informationen bzgl. unabhängiger Handlungsweisen, Anschauungen und Eigenschaften zu gewinnen, wird die Gruppendiskussion angewandt, um nach Möglichkeit eine detaillierte Übersicht über Anschauungen und Gedanken unterschiedlicher Patientinnen und Patienten hervorzubringen.²⁴⁵

Die Gruppendiskussion eignet sich zur Vorbereitung von standardisierten Patientenbefragungen unter zweierlei Gesichtspunkten. Einerseits kommt es vor, dass sich patientenspezifische Daten nicht direkt erheben lassen, andererseits dient die Befragung von Fokusgruppen als explorative Forschungsmethode, welche Sachverhalte aufdecken kann, die vor Untersuchungsbeginn nicht oder nur wenig bekannt waren.²⁴⁶ Auch in finanzieller Hinsicht beinhaltet die Gruppen-

befragung gegenüber standardisierten Einzelbefragungsmethoden Kosteneinsparungspotentiale und bietet insbesondere im Bereich der Patientenzufriedenheitsmessung oftmals das einzige effizient einsetzbare Befragungsinstrument zur Erhebung von Zufriedenheitskriterien spezieller Patientengruppen, wie bspw. Personen mit schweren physischen oder psychischen Einschränkungen.²⁴⁷

Die Untersuchungsmethode der Gruppendiskussion kann auch unter alternativen Gesichtspunkten vorteilhaft wirken. So ist sie mit relativ wenig Zeitaufwand verbunden, was insbesondere im direkten Vergleich mit eher längerfristigen Befragungsmethoden einen Nutzeneffekt darstellt. Darüber hinaus können durch die relativ kurze Planungs- und Durchführungsdauer kurzfristig Daten erhoben werden, welche bei der quantitativen Untersuchung vernachlässigt wurden.²⁴⁸

Das Niveau von Gruppendiskussionen ist von der Planung abhängig. So muss bereits im Vorfeld entschieden werden, welche Personengruppen teilnehmen und wie diese zur Teilnahme animiert werden sollen, wie die Diskussion in Bezug auf räumliche Gegebenheiten, Motivation und Moderation realisiert werden soll, welche Fragen geklärt werden sollen und wie Transkription, Analyse und die Darstellung der Ergebnisse von statten gehen sollen. Der Planungsprozess sollte dabei stets den Datenanforderungen entsprechen, eine Vergleichsmöglichkeit bieten, ethische Aspekte beinhalten und unter Miteinbeziehung möglicher Auswirkungen auf die DiskussionsteilnehmerInnen gestaltet werden.²⁴⁹

Im Falle einer traditionellen Gruppenbefragung sollte sich die Gruppengröße zwischen 7 und 10 potentiellen, aktuellen oder ehemaligen Patientinnen und Patienten bewegen, wobei stets darauf zu achten ist, dass die DiskussionsteilnehmerInnen sich gegenseitig sehen und hören und der Diskussion folgen können. Das Gespräch selbst soll ca. 60 Minuten dauern. Sollte der verbale Austausch jedoch, veranlasst durch Erschöpfung der TeilnehmerInnen oftmals ins Stocken geraten, empfiehlt es sich, die Teilnehmeranzahl zu reduzieren.²⁵⁰ Es gibt aber auch andere Ansätze von Gruppendiskussionen, wie bspw. Minigruppen oder Onlinediskussionen, welche nicht dem traditionellen Schema folgen. Minigruppenbefragungen sind bewusst in ihrer Teilnehmerzahl begrenzt, um den DiskussionsteilnehmerInnen mehr Zeit für ihre Ausführungen zur Verfügung zu stellen und so tiefergehende Informationen über ihre persönlichen Erlebnisse und Einstellungen zu erhalten. Im Gegensatz dazu können mit Hilfe von Onlinegruppendiskussionen TeilnehmerInnen miteinander in Verbindung gebracht werden, welche innerhalb einer traditionellen Gruppendiskussion aufgrund von temporalen oder räumlichen Faktoren nicht aufeinandertreffen würden.²⁵¹

Gruppendiskussionen eignen sich insbesondere dazu, Erkenntnisse über einen Untersuchungsbereich zu erlangen,

²⁴⁰Vgl. Magerhans (2016), S. 170.

²⁴¹Vgl. Applebaum et al. (2004), S. 78.

²⁴²Vgl. Magerhans (2016), S. 170.

²⁴³Vgl. Joo et al. (2017), S. 173.

²⁴⁴Vgl. Vogl (2014), S. 581.

²⁴⁵Vgl. Homburg und Krohmer (2006), S. 265.

²⁴⁶Vgl. Pirk und O. (2012), S. 231.

²⁴⁷Vgl. Applebaum et al. (2004), S. 44.

²⁴⁸Vgl. Pirk und O. (2012), S. 231 f.

²⁴⁹Vgl. Vogl (2014), S. 583.

²⁵⁰Vgl. Applebaum et al. (2004), S. 44.

²⁵¹Vgl. Kühn und Koschel (2011), S. 276 ff.

über welchen relativ wenig bekannt ist oder Annahmen über bestimmte Forschungsgebiete auf ihre zeitliche Konstanz hin zu überprüfen. Baker hat bspw. 1991 in einem mehrstufigen Verfahren auch die Gruppendiskussion angewandt, um ein Instrument zur Erhebung der Patientenzufriedenheit mit der hausärztlichen Behandlung zu entwickeln. Hierbei wurden zuerst unterschiedliche Zufriedenheitsfaktoren mittels Literaturrecherche ermittelt und jene Faktoren dann mittels Gruppendiskussion bewertet bzw. ergänzt. Aus den dadurch definierten Faktoren wurde ein Fragebogen erstellt, welcher nach mehreren Testdurchläufen als valides Erhebungsinstrument bestätigt und fortan als Consultation Satisfaction Questionnaire (CSQ) bezeichnet wurde.²⁵² Baker entwickelte in seiner Untersuchung aber nicht nur ein Erhebungsinstrument, sondern konnte auch darstellen, wie wichtig zwischenmenschliche Aspekte wie Respekt und Empathie oder gegenseitiges Vertrauen innerhalb der Zufriedenheitsbewertung mit der behandelnden Hausärztin / dem behandelnden Hausarzt sind.²⁵³

5.1.3. Critical Incident Technique

Die Critical Incident Technique (CIT) ist eine Untersuchungsmethode, welche unter anderem angewandt wird, um Erwartungen und Problemfelder von DienstleistungsempfängerInnen zu identifizieren.²⁵⁴ Sie setzt sich aus einer Reihe unterschiedlicher Prozessschritte zusammen und versucht, durch die Beobachtung des menschlichen Verhaltens, Lösungsansätze für praktische Probleme zu liefern, wobei als Problemansatz kritische Ereignisse im Fokus stehen.²⁵⁵ Generell kann postuliert werden, dass die CIT ein Instrument zur Erfassung von Exempeln außerordentlich effizienter oder ineffizienter Verhaltensweisen in vorweg eindeutig festgelegten Sachverhalten darstellt.²⁵⁶

In der Regel wird die Methode in Form eines Interviews angewandt, wobei die Auskunftspersonen gebeten werden, sich zu exponentiell negativen oder positiven Erfahrungen, welche sie im Rahmen des jeweiligen Themenbereiches gemacht haben, zu äußern.²⁵⁷ Danach wird die Auswertung der gesammelten Daten in mehreren Schritten durchgeführt, wobei an erster Stelle diejenigen kritischen Ereignisse in die Begutachtung miteinbezogen werden, welche einen ausreichenden Detailgenauigkeitsgrad innehaben, um die Situation, welche Zufriedenheit bzw. Unzufriedenheit hervorgerufen hat, transparent zu reflektieren. In weiterer Folge werden die unterschiedlichen kritischen Ereignisse verschiedenen Rubriken zugewiesen, wodurch sich einflussnehmende Faktoren, welche große Auswirkung auf die subjektiv empfundene Zufriedenheit ausüben, innerhalb von Überkategorien gruppieren lassen.²⁵⁸

Der Vorteil der CIT besteht vor allem darin, dass die Ereignisse, an welche sich die Befragten erinnern können, in der Regel einen sehr hohen Stellenwert in deren Wahrnehmung einnehmen und damit überwiegend strukturelle Qualitätskriterien darstellen. In diesem Zusammenhang ist jedoch auch zu beachten, dass die CIT, anlässlich der Konzentration auf außergewöhnlich positive oder negative Situationen, die umfassende Qualitätssicht nicht ausreichend darstellt.²⁵⁹

Im Bereich der Messung der Patientenzufriedenheit mit Hausärztinnen und -ärzten wird die CIT unter anderem eingesetzt, um die Qualitätskriterien bzw. Qualitätsanforderungen bestimmter Patientengruppen zu erforschen. Als Beispiel wäre hier die Studie von Marcinowicz et al. zu nennen, bei welcher explizit das von älteren Patientinnen und Patienten erwartete Ärztinnen- und Ärzterverhalten in Polen untersucht wurde. Mit Hilfe der CIT konnten insgesamt 122 qualitative Patientenkommentare protokolliert werden, wobei der Großteil der genannten Ereignisse positive Sachverhalte beschrieb. Durch die Interviews wurde festgestellt, dass für ältere Patientinnen und Patienten ausgeprägte soziale und emotionale Verhaltensweisen von Ärztinnen und Ärzten deutlich wichtiger sind als fachliche Kompetenzen.²⁶⁰

5.1.4. Beschwerde- und Vorschlagssysteme

Beschwerde- und Vorschlagssysteme sind mittels Formular, Telefonhotline oder mittels Onlineunterstützung eingerichtete Anlaufstellen, welche vor dem Hintergrund der Informationsgewinnung betrieben werden und die Möglichkeit bieten, Problembereiche schnell zu erkennen und gegebenenfalls rechtzeitig gegenzusteuern.²⁶¹ Auf Grundlage der kritischen Auseinandersetzung mit lobenden Worten und Beanstandungen bedeutsamer Anspruchsgruppen erlaubt bspw. die Analyse von Beschwerde- und Vorschlagssystemen die Kategorisierung von Eigenschaften als Grundanforderung, Leistungsanforderung, Begeisterungseigenschaft oder wertfreier Merkmalsausprägung.²⁶²

Der Nutzeneffekt dieser Methodik ergibt sich daraus, dass die Datensammlung wirtschaftlich kostengünstig durchführbar ist und Patientinnen und Patienten eine relativ einfache und anonyme Möglichkeit erhalten, ihre Präferenzen zu offenbaren. Negative Aspekte hingegen begründen sich in inexisterter Repräsentativität, wodurch sich die Evaluation von Kategorie und Ausmaß der Anmerkungen als diffizil darstellt.²⁶³

Mahler et al. haben bspw. mit Hilfe von teilstrukturierter Feedbackformularen die Zufriedenheit mit der Arzt-Patienten-Kommunikation und deren Auswirkung auf die generelle Patientenzufriedenheit untersucht. Hierfür wurden im Rahmen des HeiCare-Projektes Feedbackformulare für Personen mit chronischen Krankheiten erstellt. Gemessen wurde nicht nur die generelle Zufriedenheit mit der

²⁵²Vgl. Baker (1990), S. 487.

²⁵³Vgl. Mead und Bower (2000), S. 1091.

²⁵⁴Vgl. Matzler und Bailom (2004), S. 267.

²⁵⁵Vgl. Flanagan (1954), S. 327.

²⁵⁶Vgl. Hentschel (1992), S. 164.

²⁵⁷Vgl. Chell (2004), S. 48.

²⁵⁸Vgl. Matzler und Bailom (2004), S. 277.

²⁵⁹Vgl. Homburg und Krohmer (2006), S. 988.

²⁶⁰Vgl. Marcinowicz et al. (2014), S. 361 ff.

²⁶¹Vgl. Schimanofsky (2013), S. 186 f.

²⁶²Vgl. Hölzling (2008), S. 66.

²⁶³Vgl. Applebaum et al. (2004), S. 89.

Arzt-Patienten-Kommunikation, sondern vor allem die Beratungsqualität in Bezug auf die dauerhafte Medikamenteneinnahme, da diese insbesondere bei jener Patientengruppe von erheblicher Bedeutung ist. Danach wurden den einzelnen Ärztinnen und Ärzten die gesammelten Ergebnisse mitgeteilt und mit einer zweiten Fragerunde begonnen. Die gesamte Arbeit zielte darauf ab zu untersuchen, ob die Feedbackkommunikation bzgl. der Gesamtinformation über Arzneimittel die Patientenzufriedenheit mit Hausärztinnen und -ärzten verbessert oder nicht. Innerhalb der Untersuchung konnte eine signifikante Erhöhung der Patientenzufriedenheit nach Mitteilung der gesammelten Feedbackergebnisse an die behandelnde Hausärztin / den behandelnden Hausarzt nachgewiesen werden.²⁶⁴

5.1.5. Standardisierte Befragungsarten

Im Bereich der standardisierten Befragungsarten zur Erhebung der empfundenen Patientenzufriedenheit kann generell zwischen standardisierter mündlicher Befragung, standardisierter schriftlicher Befragung, standardisierter Telefonischer Befragung und Onlinebefragung unterschieden werden.²⁶⁵

Die standardisierte mündliche Befragung ist die anspruchsvollste der dargestellten quantitativen Befragungsarten. Der bedeutendste Vorteil dieser Befragungsform zeigt sich in der Güte der gesammelten Informationen. Die hohe Qualität resultiert daraus, dass die Interviewrahmenbedingungen beeinflusst werden können und die Befragung aktiv, bspw. durch Erläuterung komplizierter Sachverhalte, gestaltet werden kann.²⁶⁶ Da der Einfluss der Interviewerin / des Interviewers bei dieser Befragungsform relativ hoch ist, können dadurch jedoch die Ergebnisse verzerrt und möglicherweise der Anspruch auf Objektivität nicht erfüllt werden.²⁶⁷ Trotz der genannten Defizite sind standardisierte mündliche Befragungen gut geeignet, um Erkenntnisse in Bezug auf die Zufriedenheitsbeurteilung zu gewinnen.²⁶⁸

Ein Beispiel für eine Untersuchung bei der unter anderem das Instrument der normierten mündlichen Befragung zum Einsatz gekommen ist, ist die Studie von Stewart et al. aus dem Jahr 2000. Die Forscher untersuchten damals den Zusammenhang von patientenzentrierter Kommunikation innerhalb des ärztlichen Beratungsgesprächs und nachfolgendem Patientenwohlfinden sowie damit einhergehender Patientenzufriedenheit. Zu diesem Zweck wurden 39 Hausärztinnen und -ärzte zufällig ausgewählt und insgesamt 315 Patientinnen und Patienten befragt. Durch die Untersuchung konnte gezeigt werden, dass eine positive Arzt-Patienten-Kommunikation erheblichen Einfluss auf die emotionale Wahrnehmung der Patientinnen und Patienten hat, damit einen elementaren Bestandteil der Patientenzufriedenheit

darstellt und sogar zu elaborierten Behandlungsergebnissen führt.²⁶⁹

Die normierte schriftliche Befragung gleicht den wesentlichsten negativen Aspekt des normierten mündlichen Interviews – die Einflussnahme durch die Interviewerin / den Interviewer – aus. Durch das Fehlen einer aktiven Interviewerin / eines aktiven Interviewers kann jedoch die Rücklaufquote des Befragungsinstrumentes gering und wenig repräsentativ ausfallen. Darüber hinaus gibt es bei dieser Befragungsmethode keine Möglichkeit der Erläuterung komplizierter Sachverhalte oder der Nutzung von Verständnisfragen.²⁷⁰

Im Vergleich mit anderen Arten der Primärdatenerhebung ist also bei Verwendung von standardisierten schriftlichen Befragungen generell mit eher geringen Rücklaufquoten zu rechnen. Im Gegenzug bedeutet die Fragebogenaufbereitung für die Forscherin / den Forscher etwas weniger Aufwand. Dieser Umstand begründet sich darin, dass die Fragenreihenfolge nicht zwingend das ausschlaggebende Qualitätskriterium darstellt, da die antwortende Person in der Regel ausreichend Zeit hat, sich mit den Fragen zu befassen. Das bedeutet aber nicht, dass die Fragenanordnung vollends vernachlässigt werden kann. Die Bündelung von Themenbereichen und der logische Fragebogenaufbau sollten unter allen Umständen gewährt werden.²⁷¹

Als häufig genanntes Beispiel für eine standardisierte schriftliche Befragung im Patientenzufriedenheitsbereich ist die Studie von Williams und Calnan aus dem Jahr 1991 zu nennen. Mittels postalisch versandtem Fragebogen wurde versucht zu erforschen, welche Faktoren die Zufriedenheit von südostenglischen Hausarztpatientinnen und -patienten beeinflussen. Während die Gesamtzufriedenheit relativ hoch bewertet wurde, wurden die einzelnen Faktoren doch sehr unterschiedlich bewertet. Als ausschlaggebendste Zufriedenheitsfaktoren konnten die Dimensionen Kommunikation, die Arzt-Patienten-Beziehung sowie die Professionalität der behandelnden Ärztin / des behandelnden Arztes identifiziert werden.²⁷²

Das standardisierte telefonische Interview erlaubt es, Untersuchungen in geografisch relativ weit auseinanderliegenden Gebieten wirtschaftlich durchzuführen, jedoch verbindet diese Befragungsform die Nachteile des normierten mündlichen Interviews und der normierten schriftlichen Befragung. Das bedeutet, dass die Repräsentativität der Stichprobe durch die relative Anonymität von TelefonbesitzerInnen abnimmt und zeitgleich der Einfluss auf die Auskunftsperson durch das persönliche Gegenüber zunimmt.²⁷³

Darüber hinaus ist es nicht möglich, Personen oder Personengruppen ohne Telefon bzw. mit unbekannter Telefonnummer zu erreichen, verglichen mit anderen Befragungsformen ist die Interviewzeitspanne und damit der Fragenumfang eher gering und komplizierte Sachverhalte, für deren

²⁶⁴Vgl. Mahler et al. (2012), S. 178.

²⁶⁵Vgl. Homburg (2017), S. 271.

²⁶⁶Vgl. Olbrich et al. (2012), S. 75.

²⁶⁷Vgl. Wüthrich-Schneider (2000b), S. 1118.

²⁶⁸Vgl. Applebaum et al. (2004), S. 84.

²⁶⁹Vgl. Stewart et al. (2000), S. 796.

²⁷⁰Vgl. Berekoven et al. (2007), S. 110.

²⁷¹Vgl. Porst (2001), S. 1 ff.

²⁷²Vgl. Williams und Calnan (1991), S. 237.

²⁷³Vgl. Kuß (2012), S. 126.

Klärung die visuelle, auditive oder olfaktorische Wahrnehmung obligat wären, können telefonisch nur schwer geklärt bzw. erläutert werden. Trotz der genannten Nachteile überwiegen, vor dem Hintergrund temporaler oder wirtschaftlicher Aspekte, in mancher Hinsicht auch die Vorteile dieser Befragungsform. So sind standardisierte telefonische Befragungen im Vergleich zu anderen Befragungsformen relativ schnell und kurzfristig durchführbar, die Kosten sind relativ gering. Da die InterviewerInnen nicht an Öffnungszeiten gebunden sind, ist es auch möglich, relativ schwer erreichbare Personengruppen zu kontaktieren und die zu befragenden Personen können bereits im Vorhinein, bspw. per mobiler Textnachricht, über den bevorstehenden Anruf und die Befragung informiert werden.²⁷⁴

Als Beispiel soll hier wiederum die Studie von Stewart et al. aus dem Jahr 2000 herangezogen werden, welche unterschiedliche Befragungsarten verband. So wurden die Patientinnen und Patienten bereits innerhalb der Arztpraxis, jedoch noch vor ihrem Arztbesuch ausgewählt. Danach wurde das Behandlungsgespräch aufgezeichnet und nachfolgend das standardisierte Interview geführt. Abschließend jedoch wurden die Patientinnen und Patienten zwei Monate nach ihrem Praxisbesuch telefonisch kontaktiert, um ihre Erfahrungen mittels standardisierter telefonischer Befragung Review passieren zu lassen.²⁷⁵

Die Menge an Studien und Untersuchungen, welche durch Unterstützung des Internets durchgeführt worden sind, ist in den vergangenen Jahren stetig gewachsen. Dieser Umstand beruht nicht nur auf den umfassenden Vorteilen von Onlineumfragen, sondern auch darauf, dass immer mehr Daten online abrufbar sind und mittlerweile auch viele Personengruppen über das Internet erreichbar sind.²⁷⁶

Darüber hinaus vereint die Onlinebefragung die Vorteile der anderen Befragungsarten – ohne deren Nachteile vollends zu tragen. So ist es möglich, mittels Onlineumfrage geografisch weit auseinanderliegende Gebiete zu untersuchen, bestimmte vordefinierte Zielgruppen anzusprechen und mittels Video- oder Audiounterstützung komplizierte Sachverhalte abzufragen. Insgesamt verfügen Onlineumfragen häufig über eine sehr hohe Datenqualität, da die gesammelten Daten beinahe vollautomatisiert verarbeitet werden können und die TeilnehmerInnen den Erhebungszeitpunkt frei wählen können. Trotz der genannten Stärken von Onlineumfragen enthalten diese naturgemäß auch Schwächen. So können bspw. bestimmte Zielgruppen nicht oder nur schwer erreicht werden oder die Identität der TeilnehmerInnen kann nur schwer überprüft werden, was die Gefahr von Mehrfachbeantwortungen erhöht.²⁷⁷

Ungeachtet der genannten Nachteile entscheiden sich viele AnwenderInnen, veranlasst durch die stetig wachsende Fülle an Umfragesoftware und die damit einhergehende Bedienungs- und Analysequalitätsverbesserung, für die Onli-

nebefragung. Insbesondere die ökonomischen Vorteile dieser Befragungsart, durch welche relativ große Stichproben in relativ kurzer Zeit erhoben werden können, kennzeichnet für eine beachtliche Anzahl an NutzerInnen das ausschlaggebende Merkmal.²⁷⁸

Ein Beispiel für eine Onlinebefragung ist die UK National GP Patient Survey, welche die Zufriedenheit britischer Patientinnen und Patienten mit ihren jeweiligen Hausärztinnen und -ärzten widerspiegeln soll. Die Umfrage, welche seit 2006 durchgeführt wird, zählt zu den größten Umfrageprogrammen in Großbritannien, bei welchem jährlich fünf Millionen Personen zur Teilnahme eingeladen werden.²⁷⁹ Da die jährlichen Ergebnisse online öffentlich eingesehen werden können, hat sich die UK National GP Patient Survey auch zu einer häufig verwendeten Datenbasis in der Forschung entwickelt. So haben bspw. Croker et al. (2013) mit Hilfe der gesammelten Daten aus dem Jahr 2013 eine Analyse durchgeführt, um herauszufinden, welche Faktoren das Patientenvertrauen in ihre jeweiligen Hausärztinnen und -ärzte beeinflussen.²⁸⁰

5.2. Beobachtung

Mit der Untersuchungsmethode der Befragung können vor allem jene Blickwinkel und Faktoren erhoben werden, welche die Befragten persönlich wahrnehmen und formulieren können. Unbewusste Motive und Sachverhalte lassen sich demzufolge kaum durch direkte oder indirekte Befragungsformen in Erfahrung bringen. Um auch die unbewussten Sachverhalte messbar zu gestalten, hat sich die Erhebungsform der Beobachtung etabliert. Der essentielle Vorteil hierbei liegt darin, dass die Befragten nicht über die Fähigkeit verfügen müssen, die untersuchten Blickwinkel und Faktoren persönlich wahrzunehmen und zu formulieren, da die Beobachtung ohnehin unbewusste Faktoren zu untersuchen versucht.²⁸¹

Im direkten Vergleich mit der Befragung wird bei der Beobachtung die Handlungsweise der beobachteten Personen unverzüglich und direkt erfasst. Ungeachtet dessen können die hinter der Handlungsweise stehenden Motive nicht oder nur schwer direkt erfasst werden und müssen in der Regel im Nachhinein von der Forscherin / vom Forscher ausgelegt werden.²⁸² Jeder Sachverhalt, der sichtbar in Form von Prozessen oder Handlungsschemen erfasst werden kann, kann auch beobachtet werden. Mit Ausnahme des Spezialfalls der Selbstbeobachtung geschieht dies durch Unterstützung von autonomen dritten Personen. Die Beobachtung wird also durch Individuen, autonom oder unterstützend durch dementsprechende technische Unterstützung durchgeführt. Die Subjektbeobachtung kennzeichnet sich insbesondere dadurch, dass das Resultat nicht von der persönlichen Wahrnehmung der beobachteten Personen, sondern von deren Bereitwilligkeit an der Beobachtung teilzunehmen, abhängt.²⁸³

²⁷⁴Vgl. Hüfken (2014), S. 634 ff.

²⁷⁵Vgl. Stewart et al. (2000), S. 798.

²⁷⁶Vgl. Thielsch und Weltzin (2009), S. 69.

²⁷⁷Vgl. Theobald (2017), S. 25.

²⁷⁸Vgl. Thielsch und Weltzin (2009), S. 69.

²⁷⁹Vgl. Campbell et al. (2009), S. 1.

²⁸⁰Vgl. Croker et al. (2013), S. 1.

²⁸¹Vgl. Raab et al. (2009), S. 127 f.

²⁸²Vgl. Homburg und Krohmer (2006), S. 274.

²⁸³Vgl. Berekoven et al. (2007), S. 141.

In Untersuchungsbereichen, in welchen die Probanden entweder nicht auskunftswillig oder nicht auskunftsfähig sind, beschreibt die Beobachtung also die einzige Möglichkeit der Datengenerierung. Darüber hinaus weist die Beobachtung gegenüber anderen Erhebungsmethoden eine Reihe von Vorteilen auf. So können durch diese Methode bspw. in bestimmten Fällen Kosteneinsparungspotentiale erzielt werden und die Beobachterin / der Beobachter übt keinerlei Einfluss auf die Probanden aus, wodurch Verzerrungseffekte auf ein Minimum reduziert werden.²⁸⁴

Naturgemäß verfügt die Beobachtung auch über eine Reihe von Einschränkungen in Bezug auf den Zugang, etwaige Beeinträchtigungen, die temporalen Gegebenheiten oder Erhebungspotentiale.²⁸⁵ Diese Einschränkungen äußern sich in erster Linie darin, dass die beobachteten Personen sich nicht mehr natürlich verhalten, sobald sie die Beobachtungssituation erkennen, andererseits jedoch die Beobachtung von Personen, welche nicht realisieren, dass sie beobachtet werden, ethisch und rechtlich bedenklich ist. Weitere Problemfelder bestehen darin, dass Beobachtungen nicht präzise wiederholt werden können und die Beobachterin / der Beobachter Verhaltensweisen falsch interpretiert oder wahrnimmt oder ihre / seine persönliche Meinung einfließen lässt.²⁸⁶ Bei der Beobachtung wird, je nach Untersuchungsgegenstand, mehr oder minder strukturiert vorgegangen.²⁸⁷ Trotz der Vielzahl an Handlungs- und Verwendungsmöglichkeiten bestehen auch bei Beobachtungen eine Reihe von Grundanforderungen für die Sicherstellung ausreichender Repräsentativität. So sollen wissenschaftliche Beobachtungen dadurch gekennzeichnet sein, dass ein vordefiniertes Forschungsziel besteht und Hypothesen überprüft werden, ausgesuchte Faktoren sollen strukturiert für die Untersuchung ausgewählt werden, die gewonnenen Informationen sollen einer Auswertung unterzogen werden und der Beobachtungsprozess, jedoch nicht die spezifische Beobachtung, muss wiederholbar und subjektunabhängig plausibel sein, um das Gütekriterium der Objektivität zu erfüllen.²⁸⁸ Beobachtungen haben in jüngster Vergangenheit als Erhebungsmethode stetig zugenommen. Dieser Umstand ist wohl darauf zurückzuführen, dass mit Hilfe der Beobachtung Handlungsweisen untersucht werden können, welche nicht verbal hinterfragt werden können. Dadurch erhält die Beobachterin / der Beobachter umfassende Einsichten und kann unter Umständen Beweggründe für bestimmte Verhaltensweisen besser verstehen.²⁸⁹

Robbins et al. haben bspw. bereits 1993 mittels Beobachtung den Zusammenhang zwischen dem Behandlungsverhalten von Hausärztinnen und -ärzten und Patientenzufriedenheit untersucht. Zu diesem Zweck wählten sie im ersten Schritt ein Messinstrument, den Davis Observation Code (DOC), welches es erlaubt, das Verhalten von Ärztinnen und Ärzten in analysierbaren Einheiten abzubilden und

dadurch eine direkte Verbindung zur Patientenzufriedenheit herzustellen. Danach wurden 100 Patientinnen und Patienten vor Behandlungsbeginn zu ihrer Gesamtzufriedenheit mit der medizinischen Versorgung befragt. Das nachfolgende Behandlungsgespräch wurde per Videoaufnahme beobachtet und die Verhaltensweise der behandelnden Ärztin / des behandelnden Arztes den Kategorien des DOC zugeordnet. Nach der Behandlung wurden die Patientinnen und Patienten wieder zu ihrer Zufriedenheit befragt, diesmal jedoch zur Zufriedenheit mit dem spezifischen Arztbesuch. Die Auswertung der gesammelten Ergebnisse hat gezeigt, dass Gespräche über die medizinische Vorgeschichte der jeweiligen Patientinnen / Patienten sich negativ auf die Patientenzufriedenheit auswirken. Im Gegensatz dazu wurden Gespräche, welche sich gezielt mit dem derzeitigen Gesundheitsproblem befassten, positiv wahrgenommen und führten zu steigender Patientenzufriedenheit.²⁹⁰

Ein weiteres Beispiel für das Erhebungsinstrument der Beobachtung im Bereich der Patientenzufriedenheitsmessung ist die Studie von Gross et al. aus dem Jahr 1998, in welcher untersucht wurde, inwiefern sich die Behandlungsdauer und das Behandlungsgespräch auf die Patientenzufriedenheit auswirkt. Hierfür wurden 2315 Patientinnen und Patienten während ihres Praxisbesuches beobachtet und danach mittels Fragebogen zu ihrer Zufriedenheit befragt. Die Auswertung der gesammelten Ergebnisse hat gezeigt, dass Behandlungsgespräche, welche länger als 15 Minuten dauerten, sich jedenfalls positiv auf die Patientenzufriedenheit auswirkten und dass die empfundene Zufriedenheit auch durch kurze private Gespräche während der Behandlung erhöht wurde. Des Weiteren beeinflussten Faktoren, wie das Alter der Patientin / des Patienten oder der Gesundheitsstatus, ebenfalls das Zufriedenheitslevel. Innerhalb der Analyse wurde also festgestellt, dass Ärztinnen und Ärzte die Zufriedenheit ihrer Patientinnen und Patienten durch längere Behandlungen und kurze Gespräche ohne medizinischen Inhalt verbessern können.²⁹¹

Eine aktuellere Untersuchung im Zusammenhang mit Beobachtung in der Patientenzufriedenheitsforschung ist jene von McKinstry et al. aus dem Jahr 2010. Hier wurden sowohl die Unterschiede in der Beratungsqualität zwischen persönlichem und telefonischem Gesundheitsgespräch als auch die Unterschiede hinsichtlich der Patientenzufriedenheit bei Verwendung der beiden Gesundheitsberatungsmethoden untersucht. Um die Unterschiede innerhalb der beiden Methoden zu erforschen, wurden 106 Audioaufnahmen von telefonischen und persönlichen Arzt-Patienten-Gesprächen mit Hilfe drei etablierter Skalen analysiert. Die Patientenzufriedenheit hingegen wurde durch ein eigens entwickeltes und validiertes Erhebungsinstrument gemessen. Die gesammelten Ergebnisse deuten darauf hin, dass telefonische Beratungsgespräche zwar, vor allem bedingt durch den Komfortfaktor, von Patientinnen und Patienten ähnlich zufriedenstellend wahrgenommen werden wie persönliche Arzt-Patienten-Gespräche,

²⁸⁴Vgl. Homburg und Krohmer (2006), S. 274.

²⁸⁵Vgl. Berekoven et al. (2007), S. 141.

²⁸⁶Vgl. Homburg und Krohmer (2006), S. 274.

²⁸⁷Vgl. Raab et al. (2009), S. 128.

²⁸⁸Vgl. Greve und Wentura (1997), S. 13.

²⁸⁹Vgl. Berekoven et al. (2007), S. 145.

²⁹⁰Vgl. Robbins et al. (1993), S. 17.

²⁹¹Vgl. Gross et al. (1998), S. 133 f.

jedoch wird durch telefonische Gespräche die Diagnosefähigkeit und damit einhergehend die Patientensicherheit gefährdet.²⁹²

5.3. Experimentelle Forschungsdesigns

Mit Hilfe von experimentellen Forschungsdesigns wird versucht, das Verhalten unterschiedlicher Gruppen von Probanden zeitgleich, simultan oder zeitverzögert zu beobachten. Bestehen Verhaltensunterschiede zwischen den untersuchten Gruppen innerhalb spezifisch messbarer Merkmale, so können diese Diskrepanzen auf jene Merkmale (Testvariablen) zurückgeführt werden. Um Abweichungen innerhalb der untersuchten Gruppen zu identifizieren, kann je nach gewähltem Forschungsdesign entweder ein Befragungsinstrument oder die Untersuchungsmethode der Beobachtung eingesetzt werden.

Wenn der erhobene Unterschied zwischen den Gruppen ausschließlich auf die spezifischen Merkmale und nicht auf Messfehler zurückzuführen ist, kann davon ausgegangen werden, dass die Unterschiede hauptsächlich durch die identifizierten Merkmale hervorgerufen worden sind.²⁹³

Experimentelle Forschungsdesigns stellen demzufolge keine separate Untersuchungsmethode dar, sondern definieren ausschließlich die Struktur des Experimentes. Die Daten selbst werden entweder durch Beobachtung oder durch Befragung oder durch eine Kombination der beiden Methoden gewonnen.²⁹⁴ Je nach gewähltem Forschungsdesign wird zwischen dem Laborexperiment und dem Feldexperiment unterschieden. Laborexperimente sind grundsätzlich dadurch gekennzeichnet, dass diese modellhaft angeordnet sind, d.h. unter künstlichen Bedingungen stattfinden, wodurch die Probanden in der Regel erkennen, dass sie an einer Untersuchung teilnehmen. Feldexperimente hingegen werden am Ort des Erkenntnisinteresses, d.h. in der für die Probanden in diesem Zusammenhang natürlichen Umgebung, durchgeführt, wodurch die Versuchspersonen üblicherweise nicht wissen, dass sie an einem Experiment beteiligt sind. Dieser Umstand lässt darauf schließen, dass Feldexperimente eher die Verhaltensweisen der Grundgesamtheit widerspiegeln, da durch diese Methode Messfehler aufgrund von unnatürlichem Verhalten von Testpersonen gering gehalten werden können. Trotzdem muss sowohl bei der Anwendung von Laborexperimenten als auch beim Einsatz von Feldexperimenten davon ausgegangen werden, dass diese verzerrende oder verfälschende Einflüsse – sogenannte Störgrößen – beinhalten.²⁹⁵

Experimente kennzeichnen sich vor allem dadurch, dass bestimmte Merkmalsausprägungen kontrolliert repräsentiert werden. Um diese Verfahrensweise sicherzustellen, müssen alle Versuchsdurchgänge die gleichen Versuchsbedingungen

beinhalten. Verändert wird jeweils nur jenes Merkmal, dessen Effekt bestimmt werden soll. Innerhalb des Experimentes wird eine vordefinierte Menge an Untersuchungsgruppen getestet. Die Unterschiede zwischen den Gruppen lassen sich feststellen, indem innerhalb jeder Untersuchungsgruppe ausschließlich eine vordefinierte Modifikation des einflussnehmenden Merkmals angewandt wird.²⁹⁶ Generell sind experimentelle Forschungsdesigns also dadurch gekennzeichnet, dass einflussnehmende Merkmale aktiv kontrolliert, Untersuchungsmerkmale beeinflusst und die daraus resultierenden Veränderungen gemessen werden.²⁹⁷

Ein Beispiel für ein durchgeführtes Experiment im Bereich der Patientenzufriedenheitsmessung ist die Studie von Longo et al. aus dem Jahr 2006, bei welcher versucht wurde, die Stärke der Patientenpräferenz in Bezug auf die gemeinsame Entscheidungsfindung im Vergleich zu anderen Patientenzufriedenheitsattributen zu messen. An dem Experiment nahmen insgesamt 747 Patientinnen und Patienten aus Südwales teil, wobei diese insgesamt vier Untersuchungsgruppen zugeordnet wurden. Das Zuordnungskriterium war das jeweilige Krankheitsbild, wobei die teilnehmenden Patientinnen und Patienten entweder unter Vorhofflimmern oder Menorrhagie oder Symptomen der Menopause oder unter Prostatahypertrophie litten. Zuerst sollten die TeilnehmerInnen zwischen der Konsultation einer Ärztin / eines Arztes mit Risikokommunikationsausbildung oder mit gemeinsamer Entscheidungsfindungsausbildung oder mit beiden Ausbildungen wählen. Danach wurden sie gebeten, einen Fragebogen zur Behandlungszufriedenheit auszufüllen. Innerhalb der Studie konnten fünf Zufriedenheitsfaktoren identifiziert werden, welche signifikant durch die Arztwahl beeinflusst wurden. Obwohl die gemeinsame Entscheidungsfindung nicht das am höchsten bewertete Attribut darstellte, haben die Ergebnisse gezeigt, dass dieser Faktor die Patientenzufriedenheit positiv beeinflussen kann und zumindest für jene Patientinnen und Patienten, welche für die gemeinsame Entscheidungsfindung votierten, einen deutlichen Mehrwert schaffen kann.²⁹⁸

5.4. Onlinere Ressourcen

Immer mehr Patientinnen und Patienten durchsuchen Onlineforen und Portale bei auftretenden Krankheitssymptomen und teilen diese auch ihrem sozialen Mediumfeld mit.²⁹⁹ Zudem nutzt eine steigende Anzahl an Patientinnen und Patienten nicht nur die klassische Mund-zu-Mund-Propaganda, um Erfahrungen, welche sie im Gesundheitswesen gemacht haben, zu verbreiten, sondern auch Ärztebewertungsportale, soziale Medien, Foren und Blogs. Diese von den Patientinnen und Patienten selbst verfassten Erfahrungsberichte bieten eine neue Dimension der Informationsgewinnung.³⁰⁰ Jene Einträge stehen natürlich nicht nur

²⁹²Vgl. McKinstry et al. (2010), S. 298.

²⁹³Vgl. Raab et al. (2009), S. 193.

²⁹⁴Vgl. Berekoven et al. (2007), S. 146.

²⁹⁵Vgl. Homburg und Krohmer (2006), S. 280 ff.

²⁹⁶Vgl. Raab et al. (2009), S. 193.

²⁹⁷Vgl. Berekoven et al. (2007), S. 146 f.

²⁹⁸Vgl. Longo et al. (2006), S. 35.

²⁹⁹Vgl. Sharma et al. (2017), S. 302.

³⁰⁰Vgl. Greaves et al. (2013), S. 1.

den VerfasserInnen selbst, sondern auch Personen, welche bspw. nach einem Umzug auf der Suche nach einer neuen Hausärztin / einem neuen Hausarzt sind, zur Verfügung. Besonders Menschen, welche sich mit einer chronischen oder bedrohlichen Erkrankung konfrontiert sehen, konsultieren häufig Onlinemedien auf der Suche nach einer geeigneten Behandlungsmethode oder einer / einem für ihr spezielles Krankheitsbild geeigneten Ärztin / Arzt.³⁰¹

In jener Hinsicht ist insbesondere auch der Bereich der Ärztebewertungsportale von erheblicher Bedeutung, da den Bewertungen und Kommentaren von Patientinnen und Patienten in der Regel ein höheres Maß an Vertrauen entgegengebracht wird als institutionellen Veröffentlichungen.³⁰² Es wird davon ausgegangen, dass nur von ca. einem Prozent der PortalnutzerInnen regelmäßige Berichte veröffentlicht werden, weitere 9 Prozent nur gelegentlich Kommentare schreiben und 90 % als reine Informationsempfänger in derartigen Foren agieren. Vor diesem Hintergrund wird deutlich, welchen Einfluss einzelne NutzerInnen in medizinischen Foren auf aktuelle oder potentielle Patientinnen und Patienten ausüben und wie wichtig es ist, die sogenannten Patient Opinion Leaders (Meinungsführer in Gesundheitsfragen) zu erkennen.³⁰³ Während die Onlinekommunikation in Gesundheitsfragen kontinuierlich zunimmt, ist noch nicht vollständig geklärt, welchen Einfluss das Such- und Kommentarverhalten von Patientinnen und Patienten auf die Arzt-Patienten-Beziehung ausübt und inwiefern dadurch die Patientenzufriedenheit beeinflusst wird. Aus diesem Grund scheint es unentbehrlich, den Einfluss von Online-Gesundheitsinformationen auf die Arzt-Patienten-Beziehung zu verstehen, um einzelne Zufriedenheitsfaktoren zu identifizieren.³⁰⁴

Auch online sind einige Datenquellen besser zur Erhebung der Patientenzufriedenheit geeignet als andere. So sind Kommentare auf Ärztebewertungsportalen und innerhalb der Patientenforen häufig relativ informationsreich, können aber oftmals die Verteilung der Grundgesamtheit nicht abbilden. Micro Blogs hingegen beinhalten grundsätzlich dieselben Defizite, sind jedoch darüber hinaus dadurch gekennzeichnet, dass sie einen relativ geringen Informationsgehalt innehaben. So sind die unterschiedlichen Onlinere Ressourcen durch unterschiedliche Anwendungsbereiche sowie Vor- und Nachteile gekennzeichnet. Tabelle 15 fasst diese beispielhaft zusammen und enthält außerdem Informationen über die jeweilige Quelle nützlicher Informationen unterschiedlicher Onlinere Ressourcen.³⁰⁵

Hervorzuheben sind hier vor allem die Bereiche soziale Medien und Ärztebewertungsportale. Der Begriff soziale Medien umfasst die Onlinekommunikation, den Transport und die Zusammenarbeit zwischen unabhängigen Netzwerken von Einzelpersonen, Organisationen und Gemeinschaften.³⁰⁶

Soziale Medien sind gekennzeichnet durch webbasierte Netzwerke, welche Einzelpersonen, Organisationen oder Gemeinschaften erlauben, ein öffentliches oder halböffentliches Profil innerhalb eines beschränkten Systems aufzubauen, sich in einer unbestimmten Art und Weise mit anderen NutzerInnen zu verbinden bzw. in Verbindung zu setzen und Inhalte mit ihnen zu teilen.³⁰⁷

Das Zentrum sozialer Medien bilden die aktive Teilnahme der NutzerInnen, die Nutzung von kollektiver Intelligenz und die Befähigung der NutzerInnen, Änderungen an bestimmten Inhalten vorzunehmen.³⁰⁸ Innerhalb sozialer Medien können Informationen relativ schnell und einfach einer großen Menge an NutzerInnen, die über dieselben oder ähnliche Interessen verfügen wie die Informationsvermittlerin / der Informationsvermittler, weitergeleitet werden.³⁰⁹ Darüber hinaus zeigt nicht zuletzt die stetig steigende Anzahl an Nutzerkonten innerhalb sozialer Medien deren elementare Bedeutung und repliziert auch die Möglichkeiten der Patientenzufriedenheitsmessung.³¹⁰

NutzerInnen sozialer Medien berichten über Ärztinnen und Ärzte, welche sie zufriedengestellt haben, verbreiten Links und teilen Beiträge jener Arztpraxen.³¹¹ Durch die Sammlung und Aggregation von Patientenberichten und dem Patientenverhalten innerhalb sozialer Medien kann sowohl zufriedenstellende als auch als nicht hinreichend empfundene medizinische Versorgung identifiziert und das Patientenzufriedenheitsniveau gemessen werden.³¹²

Im Gegensatz zu sozialen Medien sind Ärztebewertungsportale internetbasierte soziale Netzwerke, über welche Patientinnen und Patienten Inhalte, Erfahrungsberichte und persönliche Informationen freigeben können. Jene Portale definieren damit eine relativ neue Möglichkeit der öffentlichen Berichterstattung über die Leistungen von Ärztinnen und Ärzten. Dadurch haben Patientinnen und Patienten einerseits die Möglichkeit, die Behandlung und das Behandlungsumfeld selbst zu bewerten, andererseits können aber auch spezifische Informationen über niedergelassene Ärztinnen und Ärzte abgerufen werden. Dieser Umstand wird aus Sicht der Patientin / des Patienten als Vorteil wahrgenommen, weil er bereits vor Behandlungsbeginn die Wahl der Ärztin / des Arztes unter Miteinbezug der persönlichen Präferenzen ermöglicht.³¹³

Da der patientenspezifische Wissenstransfer bzgl. ärztlichen Leistungen immer häufiger über Ärztebewertungsportale durchgeführt wird, eignen sich jene Webseiten beispiellos, um mehr über die Kriterien und die Faktorstruktur der Patientenzufriedenheit zu erfahren. Darüber hinaus erlaubt die Struktur vieler Bewertungsportale, durch welche Anonymität geboten wird, sowohl in positiver als auch in negativer Hin-

³⁰¹Vgl. Köhler (2017), S. 4.

³⁰²Vgl. Burmann et al. (2012), S. 192.

³⁰³Vgl. Köhler (2017), S. 4.

³⁰⁴Vgl. Tan und Goonawardene (2017), S. 1.

³⁰⁵Vgl. Greaves et al. (2013), S. 2.

³⁰⁶Vgl. Solomon und Tuten (2013), S. 2.

³⁰⁷Vgl. Boyd und Ellison (2007), S. 211.

³⁰⁸Vgl. Kreutzer und Hinz (2010), S. 5.

³⁰⁹Vgl. Köhler (2017), S. 7 f.

³¹⁰Vgl. Greaves et al. (2013), S. 1.

³¹¹Vgl. Köhler (2017), S. 7 f.

³¹²Vgl. Greaves et al. (2013), S. 1.

³¹³Vgl. Lagu et al. (2010), S. 942.

Tabelle 15: Onlinere Ressourcen zur Erhebung der Patientenzufriedenheit; Quelle: In Anlehnung an Greaves et al. (2013), S. 2.

Ressourcentyp	Nützliche Informationen	Vorteile	Nachteile
Soziale Medien, Beispiel: Facebook	Kommentare und Reaktionen (z.B. Likes)	Hoher Nutzeranteil, hohe öffentliche Reichweite	Wenig Informationen über persönliche (intime) Erfahrungen
Ärztbewertungsportale, Beispiel: Doc Finder	Bewertungstexte und Punktbewertungen	Bewertungen beziehen sich meist direkt auf das Behandlungserlebnis	Wird nicht von allen Patientinnen bzw. Patienten genutzt (Stichprobe entspricht nicht Menge in der Grundgesamtheit)
Patientennetzwerke, Beispiel: Patientslikeme	Erfahrungsberichte	Authentische Patientenkommunikation	Durch Demografie und Interessensgruppen vorselektiert
Micro Blogs, Beispiel: Twitter	Kurzkommentare	Hoher Nutzeranteil, hohe öffentliche Reichweite	Kurze, unstrukturierte Kommentare

sicht ein überwiegend aufrichtiges Bewertungsverhalten.³¹⁴

Durch das subjektive Patientenempfinden sind Ärztebewertungen von einzelnen NutzerInnen relativ schlecht vergleichbar. In diesem Zusammenhang wird häufig auch das Bewertungsverhalten von Patientinnen und Patienten kritisch betrachtet, da laut klassischen Marketingmodellen die Wahrscheinlichkeit, dass eine unzufriedene Patientin / ein unzufriedener Patient die nicht zufriedenstellende Leistung bewertet höher ist als die Wahrscheinlichkeit, dass eine zufriedene Patientin / ein zufriedener Patient die zufriedenstellende Leistung bewertet.³¹⁵ Diese Annahme konnte jedoch durch Untersuchungen des Bewertungsverhaltens, wie bspw. jene von Lagu et al., widerlegt werden. Sowohl innerhalb der Gesamtbewertung, als auch der Schlüsselwortsuche im Bewertungstext fallen die Bewertungen eher positiv aus.³¹⁶

Die Interaktion, welche online vor sich geht, löst eine verstärkte Auseinandersetzung der Nutzerin / des Nutzers mit dem Inhalt aus, was bedeutet, dass Patientinnen und Patienten, welche online über ihre Erfahrungen berichten, auch eine starke emotionale Beziehung zu ihrer behandelnden Ärztin / ihrem behandelnden Arzt aufbauen.³¹⁷ Online-ärztebewertungen, welche von NutzerInnen erstellt werden, sind demnach nicht mit rein faktischen Erfahrungsberichten gleichzusetzen, sondern enthalten auch Informationen über emotionale Komponenten und sind daher von erheblicher Bedeutung für die Messung der Patientenzufriedenheit. Zu beachten ist jedoch, dass es sich bei Patientenerfahrungen und der Patientenzufriedenheit um relativ komplexe Konstrukte handelt, die durch unterschiedliche Dimensionen operationalisiert werden. Der Aufbau von Ärztebewertungsportalen und die vorgegebenen Bewertungskategorien verzerren demnach die Patientenmeinung und das Patientenempfinden je nach Standardisierungsgrad.³¹⁸

Als Beispiel für die strukturierte Untersuchung der Kommentarinhalte von Ärztebewertungsportalen ist hier die Stu-

die von Kadry et al. (2011) aus dem Jahr 2011 zu nennen. Innerhalb der Untersuchung wurden insgesamt 4999 Patientenkommentare von 10 verschiedenen Ärztebewertungsportalen analysiert. Durch die Studie konnte unter anderem, wie auch schon bei der Studie von Lagu et al., gezeigt werden, dass die meisten Patientinnen und Patienten ihr Behandlungserlebnis positiv bewerten. Ein weiteres Forschungsziel war die Gruppierung der untersuchten Zufriedenheitsdimensionen, welche innerhalb der Kategorien Kommunikation, Zugang, Einrichtung, Praxispersonal und Gesamtzufriedenheit klassifiziert wurden. Dadurch wurde bspw. festgestellt, dass die Kategorie-Priorisierung einzelner Patientinnen und Patienten abhängig von ihrer empfundenen Gesamtzufriedenheit ist.³¹⁹

6. Zusammenfassung und Ausblick

Die Arzt-Patienten-Beziehung ist in vielerlei Hinsicht von erheblicher Bedeutung.³²⁰ Obwohl diese durchaus Elemente von Dienstleistungsprozessen enthält, ist sie dennoch nicht vollständig mit einer materiellen Dienstleistung vergleichbar.³²¹ Ein Element, welches zweifelsohne sowohl bei materiellen Dienstleistungsprozessen als auch innerhalb der Arzt-Patienten-Beziehung an Bedeutung gewonnen hat, ist die subjektiv empfundene Kunden- bzw. Patientenzufriedenheit.³²²

Patientenzufriedenheit entsteht grundsätzlich, wenn Patientinnen und Patienten die erhaltene gesundheitliche Leistung ihrer vordefinierten Erwartung gegenüberstellen.³²³ Die Erwartungshaltung wird hierbei aufgrund von Erfahrungen,³²⁴ äußeren Einflüssen³²⁵ und persönlichen Präferenzen gebildet.³²⁶ Der Vergleichsprozess selbst kann grundsätzlich

³¹⁹Vgl. Kadry et al. (2011), o.S.

³²⁰Vgl. Hölzing (2008), S. 203.

³²¹Vgl. Bestmann und Verheyen (2010), S. 7.

³²²Vgl. Dampz et al. (2012), S. 8.

³²³Vgl. Sitzia und Wood (1997), S. 1840.

³²⁴Vgl. Piekenbrock (o.J.), o.S.

³²⁵Vgl. Lin et al. (2012), S. 55.

³²⁶Vgl. Piekenbrock (o.J.), o.S.

³¹⁴Vgl. Geierhos et al. (2015), S. 1.

³¹⁵Vgl. Ma et al. (2015), S. E 16.

³¹⁶Vgl. Lagu et al. (2010), S. 942.

³¹⁷Vgl. Köhler (2017), S. 7 f.

³¹⁸Vgl. Reimann und Strech (2010), S. 1.

drei Ausprägungen annehmen. Entspricht die erhaltene Leistung der erwarteten Leistung, löst dies generell Zufriedenheit aus.³²⁷ Im Gegensatz dazu führt die Leistungserfüllung über bzw. unter dem Vergleichsniveau zu Begeisterung oder Unzufriedenheit.³²⁸ Dieser Vergleichsprozess hat allgemeine Gültigkeit für die Bewertung von Leistungsfaktoren, gilt jedoch nicht zwingend für Zufriedenheitsfaktoren, welche als Grundvoraussetzung erwartet werden oder solche, die nicht vorausgesetzt werden.³²⁹

Um die Faktorstruktur der Patientenzufriedenheit zu kategorisieren, wurden in der Vergangenheit unterschiedliche Modelle entwickelt.³³⁰ Ein Ansatz, welcher hierbei häufig Anwendung findet, ist die Grobkategorisierung von Patientenzufriedenheit innerhalb von drei Bereichen, wobei häufig nur einer dieser Bereiche durch die Hausärztin / den Hausarzt beeinflussbare Zufriedenheitsfaktoren enthält.³³¹ Zu den beeinflussbaren Zufriedenheitsfaktoren zählen hierbei bspw. Zugang und Erreichbarkeit, Komfort, Kommunikation, wahrgenommene Behandlungsqualität, Fürsorge sowie Einrichtung und technische Ausstattung.³³²

Zur Erhebung der Ausprägungsstärke unterschiedlicher beeinflussbarer Patientenzufriedenheitsfaktoren wurden in der Vergangenheit unterschiedliche Messskalen entwickelt.³³³ Diese Skalen geben einerseits Aufschluss über die häufig untersuchten und damit bedeutsamen Zufriedenheitsdimensionen, können aber auch für eigene Zufriedenheitserhebungen als Erhebungsinstrument oder als Grundlage für die Erstellung eines Erhebungsinstrumentes herangezogen werden.³³⁴ Die in dieser Arbeit beschriebenen Skalen zur Messung der Patientenzufriedenheit enthalten insgesamt zwischen 5 und 11 unterschiedliche Zufriedenheitsdimensionen, wobei der obere Wert nach der Faktorenbündelung auf maximal 9 Zufriedenheitsfaktorkategorien korrigiert wurde.³³⁵

Bedingt durch unterschiedliche Untersuchungsschwerpunkte weisen die einzelnen Skalen untereinander Unterschiede hinsichtlich der Definition und Untersuchung einzelner Zufriedenheitsfaktoren auf. Trotz dieses Umstandes und obwohl die Arbeit Skalen aus beinahe vier Dekaden der Patientenzufriedenheitsmessungen beinhalten, sind dennoch eindeutige Übereinstimmungen zu erkennen. So werden innerhalb von allen Skalen zwischenmenschliche Aspekte und die ärztliche Fachkompetenz als einflussnehmende Zufriedenheitsfaktoren gemessen. Darüber hinaus wird die verbale Kommunikation häufig als Zufriedenheitsfaktor mit erheblichem Einfluss dargestellt.

Um die empfundene Patientenzufriedenheit zu erheben, können unterschiedliche Methoden und Instrumente einge-

setzt werden, wobei hier auch Überlegungen hinsichtlich der Erhebungsplanung und des Erhebungsdesigns zu beachten sind. In dieser Arbeit vorgestellt wurden in dieser Hinsicht das Tiefeninterview, die Gruppendiskussion, die Critical Incident Technique, die Nutzung eines Vorschlags- und Beschwerdesystems, standardisierte mündliche, schriftliche und telefonische Befragung sowie Onlinebefragung, Beobachtung, experimentelle Forschungsdesigns und die Nutzung von Onlineressourcen zur Erhebung der Patientenzufriedenheit. Die angeführten Praxisbeispiele zeigen, dass jede der vorgestellten Methoden zur Erhebung der Patientenzufriedenheit mit Hausärztinnen und -ärzten herangezogen werden kann und je nach gewähltem Forschungsschwerpunkt mehr oder weniger zielführend wirkt.

Wie innerhalb der Arbeit beschrieben wurde, handelt es sich bei Patientenzufriedenheit um kein starres Konstrukt, welches sich im Zeitverlauf nicht oder kaum verändert.³³⁶ Dieser Umstand kann einerseits darauf zurückgeführt werden, dass sich Zufriedenheitsfaktoren im Zeitverlauf innerhalb der Faktorstruktur bewegen, also bspw. von Leistungs- zu Basisfaktoren werden.³³⁷ Andererseits sind auch äußere Umwelteinflüsse dafür verantwortlich, dass sich die Bewertungskriterien unterschiedlicher Faktoren verändern bzw. neue Faktoren oder Faktorenbündel hinzukommen.³³⁸ Vor diesem Hintergrund wird deutlich, dass die Kategorisierung von Zufriedenheitsdimensionen im Allgemeinen und von Patientenzufriedenheitsdimensionen im Speziellen grundsätzlich nicht zu allgemein gültigen, vor allem aber nicht zu zeitlich unbegrenzt gültigen, Ergebnissen führen kann.

Abschließend ist festzuhalten, dass durch die gewählte Literatursuch- und Auswahlstrategie naturgemäß eine Reihe von Untersuchungen, welche möglicherweise zu anderen oder zusätzlichen Ergebnissen im Bereich der Patientenzufriedenheitsmessung geführt hätten, nicht berücksichtigt werden konnten. Dieser Umstand relativiert sich aber einerseits durch die umfassende und breit gefasste Literatursuche und -auswahl, andererseits führen die identifizierten Quellen, nach Normierung der Faktorenbündel teilweise durchaus zu vergleichbaren und ähnlichen Ergebnissen. Dieser Umstand lässt darauf schließen, dass zumindest der Kernbereich der Patientenzufriedenheitsmessung mit Hausärztinnen und -ärzten in ausreichendem Maße erfasst wurde und die Arbeit somit eine ideale Grundlage zur Durchführung eigener praktischer oder wissenschaftlicher Patientenzufriedenheitserhebungen darstellt.

³²⁷Vgl. Homburg und Krohmer (2006), S. 44.

³²⁸Vgl. Töpfer (2008a), S. 50.

³²⁹Vgl. Matzler et al. (2004), S. 19.

³³⁰Vgl. Alberto Sánchez et al. (2014), S. 153.

³³¹Vgl. Fan et al. (2005), S. 227 f.

³³²Vgl. Shelton (2000), S. 29.

³³³Vgl. Gill und White (2009), S. 11.

³³⁴Vgl. Kimberlin und Winterstein (2008), S. 2280.

³³⁵Siehe Anhang ab Seite 104.

³³⁶Vgl. Crow et al. (2002), S. 11.

³³⁷Vgl. Bartikowski (2002), S. 11.

³³⁸Vgl. Lin et al. (2012), S. 55.

Literatur

- Aboulgate, A., Abel, G., Elliott, M. N., Parker, R. A., Campbell, J., Lyratzopoulos, G., und Roland, M. Do english patients want continuity of care, and do they receive it? *British Journal of General Practice*, 62(8):567–e575, 2012.
- Adler, R., Vasiliadis, A., und Bickell, N. The relationship between continuity and patient satisfaction: a systematic review. *Family Practice*, 27(4):171–178, 2010.
- Al-Abri, R. und Al-Balushi, A. Patient satisfaction survey as a tool towards quality improvement. *Oman Medical Journal*, 29(1):3 – 7, 2014.
- Alberto Sánchez, C., Javier Prado-Galbarro, F., García-Pérez, S., und Sarría Santamera, A. Factors associated with patient satisfaction with primary care in Europe: results from the EU primecare project. *Quality in primary care*, 22(3):147 – 155, 2014.
- Aldana, J. M., Piechulek, H., und Al-Sabir, A. Client satisfaction and quality of health care in rural Bangladesh. *Bulletin of the World Health Organization*, 79(6):512–517, 2001.
- Alrubaiee, L. und Alkaa'ida, F. The mediating effect of patient satisfaction in the patients. In *Perceptions of Healthcare Quality–Patient Trust Relationship*, *International Journal of Marketing Studies*, volume 3, pages 103–127. Citeseer, 2011.
- Applebaum, R. A., Straker, J. K., und Geron, S. *Patientenzufriedenheit: Benennen, bestimmen, beurteilen*. Huber, Bern – Göttingen – Toronto – Seattle, 2004.
- Bailom, F., Tschernernjak, D., Matzler, K., und Hinterhuber, H. Durch strikte Kundennähe die Abnehmer begeistern. *Harvard Business Manager*, 20(1): 47–56, 1998.
- Baker, R. Development of a questionnaire to assess patients' satisfaction with consultations in general practice. *British Journal of General Practice*, 40 (12):487–490, 1990.
- Bartikowski, B. Kundenzufriedenheit - verfahren zur messung der indifferenzzone. 2002. Lohmar – Köln.
- Benkenstein, M. *Neue Herausforderungen an das Dienstleistungsmarketing*. Springer-Verlag, Wiesbaden, 2008.
- Berekoven, L., Eckert, W., und Ellenrieder, P. *Marktforschung: methodische Grundlagen und praktische Anwendung*. Number 12. Springer-Verlag, Wiesbaden, 2007.
- Berger, B. und Lentz, M. Bewertung deutschsprachiger Instrumente zur Erhebung von Patientenzufriedenheit. *Gesellschaft für Gesundheitsmarktanalysen*, 2006. Hamburg.
- Bertakis, K. D., Roter, D., und Putnam, S. M. The relationship of physician medical interview style to patient satisfaction. *Journal of Family Practice*, 32(2):175–182, 1991.
- Bestmann, B. und Verheyen, F. Patientenzufriedenheit. Ergebnisse einer repräsentativen Studie in der ambulanten ärztlichen Versorgung. WINEG Wissen 01, S. 1 – 22, 2010. URL <https://www.tk.de/centaurus/servlet/contentblob/224246/Datei/1014/WINEG%20Wissen%20-%20Patientenzufriedenheit.pdf>. 16.10.2017.
- Bhana, N., Thompson, L., Alchin, J., und Thompson, B. Patient expectations for chronic pain management. *Journal of primary health care*, 7(2):130–136, 2015.
- Bieber, C., Nicolai, J., Müller, K. G., und Eich, W. Der Fragebogen zur Arzt-Patient-Interaktion (FAPI)–Validierung und psychometrische Optimierung anhand einer Stichprobe chronischer Schmerzpatienten. *Diagnostik und Evaluation*, 4(1):78–93, 2011.
- Bielińska, M. Lexikographische Metatexte. 2010. Frankfurt am Main.
- Bischoff, A. *Möglichkeiten zur Erhöhung der Rücklaufquote bei schriftlichen Befragungen*. Hamburg, 2002.
- Bitzer, E. M., Dierks, M.-L., Dörning, H., und Schwartz, F.-W. Zufriedenheit in der Arztpraxis aus Patientenperspektive. *Zeitschrift für Gesundheitswissenschaften*, 7(3):196–209, 1999.
- Bjertnaes, O. A., Lyngstad, I., Malterud, K., und Garratt, A. The Norwegian EUROPEP questionnaire for patient evaluation of general practice: data quality, reliability and construct validity. *Family Practice*, 28(3):342–349, 2011.
- Bortz, J. und Döring, N. *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. Springer, Berlin – Heidelberg, 5 edition, 2016.
- Bösener, K. *Kundenzufriedenheit, Kundenbegeisterung und Kundenpreisverhalten*. Springer, Wiesbaden, 2015.
- Bower, P. und Roland, M. O. Bias in patient assessments of general practice: general practice assessment survey scores in surgery and postal responders. *British Journal of General Practice*, 53(2):126–128, 2002.
- Bower, P., Mead, N., und Roland, M. What dimensions underlie patient responses to the general practice assessment survey? a factor analytic study. *Family Practice*, 19(5):489–495, 2002.
- Bowling, A. und Ebrahim, S. Measuring patients' preferences for treatment and perceptions of risk. *Quality in Health*, 10(1):i2–i8, 2001.
- Boyd, D. M. und Ellison, N. B. Social network sites: Definition, history, and scholarship. *Journal of computer-mediated Communication*, 13(1):210–230, 2007.
- Brosius, H.-B., Haas, A., und Koschel, F. *Methoden der empirischen Kommunikationsforschung*. Springer, Heidelberg, 6 edition, 2012.
- Bruhn, M. *Qualitätsmanagement für Dienstleistungen - Grundlagen, Konzepte, Methoden*. 7 edition, 2008. Berlin Heidelberg.
- Brüsemeister, T. Qualitative forschung: Ein überblick. In 2, editor, *Qualitative Forschung*. Springer, Wiesbaden, 2008.
- Burmann, C., Halaszovich, T., und Hemmann, F. Identitätsbasierte Markenführung, 2012. Wiesbaden.
- C., W.-J. und C., Y.-H. Patient satisfaction analysis: Identifying key drivers and enhancing service quality of dental care. *Journal of Dental Sciences*, 8(9):239 – 247, 2013.
- Campbell, J. L., Smith, P., Nissen, S., Bower, P., Elliott, M., und Roland, M. The GP Patient Survey for use in primary care in the National Health Service in the UK – development and psychometric characteristics. *BMC Family Practice*, 10, S. 1 – 8, 2009. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2736918/>. 30.10.2017.
- Campbell, J., Ramsay, J., und Green, J. Age, gender, socioeconomic, and ethnic differences in patients' assessments of primary health care. *Quality in Health Care*, 10(2):90–95, 2001.
- Casalino, L., Pesko, M., Ryan, A., Mendelsohn, J., Copeland, K., Ramsay, P., X., S., Rittenhouse, D., und Shortell, S. Small primary care physician practices have low rates of preventable hospital admissions. *Health Affairs*, 33 (9), S. 1 – 9, 2014. URL <https://www.ncbi.nlm.nih.gov/pubmed/25122562>. 30.10.2017.
- Chell, E. Critical incident technique. *Cassell / Symon*, page 45 – 60, 2004.
- Cohidon, C., Cornuz, J., und Senn, N. Primary care in Switzerland: evolution of physicians' profile and activities in twenty years (1993–2012). *BMC Family Practice*, 16, S. 1 – 9, 2015. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4546123/>. 30.10.2017.
- Cousin, G., M., S., D.L., R., und Hall, J. Concordance between physician communication style and patient attitudes predicts patient satisfaction. *Patient Education and Counseling*, 87(9):193–197, 2012.
- Crocker, J., Swancutt, D., Roberts, M., Abel, G., Roland, M., und Campbell, J. Factors affecting patients' trust and confidence in GPs: evidence from the English national GP patient survey. *BMJ Open*, 3, S. 1 – 8, 2013. URL <http://bmjopen.bmj.com/content/bmjopen/3/5/e002762.full.pdf>. 30.10.2017.
- Crow, R., Gage, H., Hampson, S., Hart, J., Kimber, A., und Storey, L. The measurement of satisfaction with healthcare: implications for practice from a systematic review of the literature. *Health Technology Assessment*, 6(1): 1–244, 2002.
- Dagger, T., Sweeney, J., und Johnson, L. A hierarchical model of health service quality: Scale development and investigation of an integrated model. *Journal of Service Research*, 10(2):123–142, 2007.
- Dampz, M., Falcato, L., und Beck, T. Patientenzufriedenheit – Leuchtturm oder Irrlicht auf dem Weg zur Qualitätsbeurteilung ambulanter Suchtbehandlungen. *Abhängigkeiten*, 17(1):7–29, 2012.
- Derksen, F., Bensing, J., und Lagro-Janssen, A. Effectiveness of empathy in general practice: a systematic review. *British Journal of General Practice*, 63(1):76–84, 2013.
- Dierks, M.-L. und W., S. F. Rollenveränderungen durch New Public Health – Vom Patienten zum Konsumenten und Bewerter von Gesundheitsdienstleistungen. *Bundesgesundheitsblatt - Gesundheitsforschung – Gesundheitsschutz*, 44(8):796–803, 2001.
- Dimova, R., Stoyanova, R., und Keskinova, D. The EUROPEP questionnaire for patient's evaluation of general practice care: Bulgarian experience. *Croatian Medical Journal*, 58(1):63–74, 2017.
- Donabedian, A. The quality of care: how can it be assessed. *Journal of the American Medical Association*, 105(12):1743–1748, 1988.
- Eide, T., Straand, J., Melbye, H., Rortveit, G., Hetlevik, I., und E., R. Patient experiences and the association with organizational factors in gene-

- ral practice: results from the Norwegian part of the international, multicentre, cross sectional QUALICOPC study. *BMC Health Services Research*, 17, S. 1 - 9, 2016. URL <https://www.ncbi.nlm.nih.gov/pubmed/27553244>. 30.10.2017.
- Fan, V., Burman, M., Mc Donnell, M., und Fihn, S. Continuity of care and other determinants of patient satisfaction with primary care. *Journal of General Internal Medicine*, 20(3):226-233, 2005.
- Fenton, J., Jerant, A., Bertakis, K., und Franks, P. The cost of satisfaction- a national study of patient satisfaction, health care utilization, expenditures, and mortality. *Archives of internal medicine*, 16(5):403-411, 2012.
- Flanagan, J. The critical incident technique. *Psychological Bulletin*, 51(4): 327-358, 1954.
- Frodl, A. Marketing im Gesundheitsbetrieb, 2011. Wiesbaden.
- Geierhos, M., Schulze, S., und Bäumer, F. Der zufriedene Patient 2.0: Analyse anonymierter Arztbewertungen im Web 2.0. Working Papers des KVF NRW, 3, S. 1 - 18, 2015. URL http://www.ssoar.info/ssoar/bitstream/handle/document/44073/ssoar-2015-geierhos_et_al-Der_zufriedene_Patient_20_Analyse.pdf?sequence=1. 30.10.2017.
- Gericke, C., Schiffhorst, G., Busse, R., und Häussler, B. Ein valides Instrument zur Messung der Patientenzufriedenheit in ambulanter haus- und fachärztlicher Behandlung: das Qualiscope-A. *Gesundheitswesen*, 66(11): 723-731, 2004.
- Gill, L. und White, L. A critical review of patient satisfaction. *Leadership in Health Services*, 22(1):8-19, 2009.
- Greaves, F., Ramirez-Cano, D., Millett, C., Darzi, A., und Donaldson, L. Harnessing the cloud of patient experience: using social media to detect poor quality healthcare. *BMJ Quality & Safety Online*, 22 (3), S. 1 - 4, 2013. URL <http://qualitysafety.bmj.com/content/22/3/251.long>. 30.10.2017.
- Greve, W. und Wentura, D. Wissenschaftliche Beobachtung, 1997. Weinheim.
- Grol, R., Wensing, M., Mainz, J., Ferreira, P., Hearnshaw, H., Hjortdahl, P., Olesen, F., Ribacke, M., Spensert, T., und Szécsényi, J. Patients' priorities with respect to general practice care: an international comparison. *Family Practice*, 16(1):4-11, 1999.
- Grol, R., Wensing, M., Mainz, J., Jung, H., Ferreira, P., Hearnshaw, H., Hjortdahl, P., Olesen, F., Reiss, S., Ribacke, M., und Szécsényi, J. Patients in Europe evaluate general practice care: an international comparison. *British Journal of General Practice*, 50(11):882-887, 2000.
- Gross, D., Zyzanski, S., Borawski, E., Cebul, R., und Stange, K. Patient satisfaction with time spent with their physician. *Journal of Family Practice*, 47(2):133-137, 1998.
- Ha, J. F., Anat, S., und Longnecker, N. Doctor-patient communication: A review. *The Ochsner Journal*, 10(1):38-43, 2010.
- Häder, M. Empirische Sozialforschung, 2015. Wiesbaden.
- Hagedoorn, M., Uijl, S., Van Sonderen, E., Ranchor, A., Grol, B., Otter, R., Krol, B., Van den Heuvel, W., und Sandermann, R. Structure and reliability of Ware's patient satisfaction questionnaire III: Patients' satisfaction with oncological care in the Netherlands. *Medical Care*, 41(2):254-263, 2003.
- Hekkert, K., Cihangir, S., Kleefstra, S., Van den Berg, B., und Kool, R. Patient satisfaction revisited: a multilevel approach. *Social Sciences and Medicine*, 69(1):68-75, 2009.
- Hentschel, B. Dienstleistungsqualität aus Kundensicht, 1992. Wiesbaden.
- Hogan, N., Li, H., Pezaro, C., Roberts, N., Schmidt, E., und Martin, J. Searching for a written patient feedback instrument for patient-medical student consultations. *Advances in Medical Education and Practice*, 8(10): 171-178, 2017.
- Hölzing, J. Die Kano - Theorie der Kundenzufriedenheitsmessung, 2008. Wiesbaden.
- Homburg, C. Marketingmanagement, 6, 2017. Wiesbaden.
- Homburg, C. und Bucerius, M. Kundenzufriedenheit als Managementherausforderung. *Homburg*, pages 53-89, 2006.
- Homburg, C. und Krohmer, H. Marketingmanagement, 2, 2006. Wiesbaden.
- Homburg, C. und Rudolph, B. Theoretische Perspektiven zur Kundenzufriedenheit. *Simon / Homburg*, pages 31-55, 1997.
- Hüfken, V. Telefonische Befragung. *Baur / Blasius*, pages 631-642, 2014.
- Hughes, S. L., Weaver, F. M., Giobbie-Hurder, A., Manheim, L., Henderson, W., Kubal, J. D., Ulasevich, A., und Cummings, J. Effectiveness of team-managed home-based primary care: a randomized multicenter trial. *Journal of the American Medical Association*, 117(22):2877-2885, 2000.
- Hulka, B., Zyzanski, S., Cassel, J., und Thompson, S. Scale for the measurement of attitudes towards physicians and primary medical care. *Medical Care*, 8(5):429-436, 1970.
- Ipsen, D. Das Konstrukt Zufriedenheit. *Soziale Welt*, 29(1):44-53, 1978.
- Jackson, J., Chamberlinc, J., und Kroenke, K. Predictors of patient satisfaction. *Social Science and Medicine*, 52(4):609-620, 2001.
- Jalil, A., Zakar, R., Zakar, M., und Fischer, F. Patient satisfaction with doctor-patient interactions: a mixed methods study among diabetes mellitus patients in Pakistan. *BMC Health Services Research*. 17 (1), S. 1 - 13, 2017. URL <https://www.ncbi.nlm.nih.gov/pubmed/28222721>. 30.10.2017.
- Janicijevic, I., Seke, K., Djokovic, A., und Filipovic, T. Healthcare workers satisfaction and patient satisfaction-where is the linkage. *Hippokratia*, 17(2):157-162, 2013.
- Joo, J., Cho, J. J., Kwon, Y. J., Lee, Y., und Shin, D. A qualitative study of satisfaction with the community-based primary care project among primary care patients and its efficacy. *Korean Medical Association*, 60(2):173-182, 2017.
- Kaba, R. und Sooriakumaran, P. The evolution of the doctor-patient relationship. *International Journal of Surgery*, 2(5):57-65, 2007.
- Kadry, B., Chu, L. F., Kadry, B., Gammass, D., und Macario, A. Analysis of 4999 online physician ratings indicates that most patients give physicians a favorable rating. *Journal of Medical Internet Research*, 13(4):1-18, 2011. 17.10.2017.
- Karner, R. Kundenorientierte Unternehmensführung: Fallbeispiel KTW Software & Consulting, Tirol. *Hinterhuber / Matzler*, pages 611-635, 2004.
- Keiner, D. Methodischer Vergleich von verschiedenen Instrumenten zur Messung der Lebensqualität bei Osteoporose-Patienten, 2006. Göttingen.
- Kick, A. Die Rolle von Patienten und Kunden- Ethische Verantwortung des Therapeuten. *Deutsches Ärzteblatt*, 103(18):A1206-A1208, 2006.
- Kim, S., Kaplowitz, S., und Johnston, M. The effects of physician empathy on patient satisfaction and compliance. *Evaluation & The Health Professions*, 27(9):237-251, 2004.
- Kimberlin, C. und Winterstein, A. Validity and reliability of measurement instruments used in research. *American Journal of Health-System Pharmacy*, 65(12):2276-2284, 2008.
- Kirchgeorg, M. Stichwort: Zufriedenheit. *Gabler Wirtschaftslexikon*, o.J. URL <http://wirtschaftslexikon.gabler.de/Archiv/11227/zufriedenheit-v6.html>. 22.05.2017.
- Köhler, M., A. / Gründer. Online-Marketing für das erfolgreiche Krankenhaus, 2017. Berlin.
- Kontopantelis, E., Roland, M., und Reeves, D. Patient experience of access to primary care: identification of predictors in a national patient survey. *BMC Family Practice*, 11:1-15, 2010. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2936332/>. 30.10.2017.
- K.P. W. Discursive versus information processing perspectives on a bioethical problem- the case of 'unrealistic' patient expectations. *Theory & Psychology*, 14(2):191 - 203, 2004.
- Kreutzer, R. und Hinz, J. Möglichkeiten und Grenzen von Social Media Marketing, 2010. Berlin.
- Kühn, T. und Koschel, K.-V. Gruppendiskussionen- ein Praxis-Handbuch, 2011. Wiesbaden.
- Kuß, A. Marktforschung: Grundlagen der Datenerhebung und Datenanalyse, 4, 2012. Wiesbaden.
- Lagu, T., Hannon, N., Rothberg, M., und Lindenauer, P. Patients' evaluations of health care providers in the era of social networking: An analysis of physician-rating websites. *Journal of General Internal Medicine*, 25(9): 942-946, 2010.
- Laidsaar-Powell, R., Bu, S., und Mc Caffery, K. Partnering with involving patients. *Martin / Di Matteo*, pages 84-108, 2014.
- Larsen, D. L., Attkisson, C. C., Hargreaves, W. A., und Nguyen, T. D. Assessment of client/patient satisfaction: development of a general scale. *Evaluation and Program Planning*, 2(3):197-207, 1979.
- Lin, M., Kan, W., und Chen, C. Patients' perceptions and expectations of family participation in the informed consent process of elective surgery in Taiwan. *Asian Nursing Research*, 6(2):55-59, 2012.
- Linde, K., Witt, C., Streng, A., Weidenhammer, W., Wagenpfeil, S., Brinkhaus, B., Willich, S., und Melchart, D. The impact of patient expectations on outcomes in four randomized controlled trials of acupuncture in patients with chronic pain. *Pain*, 128(3):264-271, 2007.
- Löffler, U. und Von der Heyde, C. Stichproben-Verfahren in der Umfragefor-

- schung, 2, 2014. Wiesbaden.
- Longo, M., Cohen, D., Hood, K., Edwards, A., Robling, M., Elwyn, G., und Russell, I. Involving patients in primary care consultations: assessing preferences using discrete choice experiments. *British Journal of General Practice*, 56(1):35–42, 2006.
- Ma, L., Kaye, A., Bean, M., Vo, N., und Ruan, X. A five-star doctor? online rating of physicians by patients in an internet driven world. *Pain Physician*, 18(1):E15–E17, 2015.
- Magerhans, A. Marktforschung- eine praxisorientierte Einführung, 2016. Wiesbaden.
- Mahler, C., Hermann, K., Jank, S., Haefeli, W., und Szecsenyi, J. Can a feedback report and training session on medication counseling for general practitioners improve patient satisfaction with information on medicines. *Patient Prefer Adherence*, 6(10):179–186, 2012.
- Mangione, C., Lee, P., Gutierrez, P., Spritzer, K., Berry, S., und Hays, R. Development of the 25-item national eye institute visual function questionnaire. *Archives of Ophthalmology*, 3(1):1050–1058, 2001.
- Marcinowicz, L., Pawlikowska, T., und Oleszczyk, M. What do older people value when they visit their general practitioner? a qualitative study. *European Journal of Ageing*, 11(4):361 – 367, 2014.
- Marshall, G. N. und Hays, R. D. The patient satisfaction questionnaire short-form (psq- 18), 1994. Santa Monica.
- Matzler, K. und Bailom, F. Messung von kundenzufriedenheit. *Hinterhuber / Matzler*, page 263 – 294, 2004.
- Matzler, K., Stahl, H., und Hinterhuber, H. Die customer-based view der unternehmung. *Hinterhuber / Matzler*, q:3 – 30, 2004.
- Mayring, P. Psychologie des glücks, 1991. Stuttgart.
- Mazor, K., Clauser, B., Field, T., Yood, R., und Gurwitz, J. A demonstration of the impact of response bias on the results of patient satisfaction surveys. *Health Service Research*, 37(5):1403 – 1417, 2002.
- McKinstry, B., Hammersley, V., Burton, C., Pinnock, H., Elton, R., Dowell, J., Sawdon, N., Heaney, D., Elwyn, G., und Sheikh, A. The quality, safety and content of telephone and face-to-face consultations: a comparative study. *Quality & Safety in Health Care*, 19(4):298 – 303, 2010.
- Mead, N. und Bower, P. Patient-centredness: a conceptual framework and review of the empirical literature. *Social Science & Medicine*, 51(7):1087 – 1110, 2000.
- Meffert, H. und Bruhn, M. Dienstleistungsmarketing Grundlagen – Konzepte – Methoden, 7, 2012. Wiesbaden.
- Mehmet, Y. Qualitätsurteile in Patientenbefragungen, 2011. Wiesbaden.
- Milano, M., Mola, E., Collecchia, G., Del Carlo, A., Giancane, R., Visentin, G., und Nicolucci, A. Validation of the italian version of the europe instrument for patient evaluation of general practice care. *The European Journal of General Practice*, 13(2):92 – 94, 2007.
- Moosbrugger, H. und Kelava, A. Testtheorie und fragebogenkonstruktion, 2, 2012. Wiesbaden.
- Nerdinger, F., Neumann, C., und Cruth, S. Kundenzufriedenheit und kundenbindung. *Moser*, pages 127 – 146, 2015.
- Neugebauer, B. und Porst, R. Patientenzufriedenheit – ein Literaturbericht. *ZUMA- Methodenbericht*, 7:1–28, 2001. URL http://www.ssoar.info/ssoar/bitstream/handle/document/48754/ssoar-2001-neugebauer_et_al-Patientenzufriedenheit_ein_Literaturbericht.pdf?sequence=1. 30.10.2017.
- Niebling, W., Unteregger, E., Böckmann, H., Dietsche, B., Mühlbacher, A., und Nübling, M. Umfassendes Qualitätsmanagement in einem Hausarzt-netz – ein Werkstattbericht. *Zeitschrift für Allgemeinmedizin*, 79(11):530 – 534, 2003.
- Nübling, M., Mühlbacher, A., und Niebling, W. Patientenbefragung in der Hausarztpraxis: Entwicklung, Validierung und Einsatz eines Instrumentes. *Zeitschrift für ärztliche Fortbildung und Qualität im Gesundheitswesen*, 98(4):301 – 308, 2004.
- Ogaji, D. Questionnaires for patient evaluation of primary health care: A systematic review and implications for the nigerian practice setting. *Journal of Community Medicine and Primary Health Care*, 29(1):18 – 34, 2017.
- Olbrich, R., Battenfeld, D., und Buhr, C. Marktforschung, 2012. Berlin – Heidelberg.
- Oliver, R. Satisfaction: A behavioral perspective on the consumer, 2010. London – New York.
- Osiya, D., Ogaji, D., und Onotai, L. Patients' satisfaction with healthcare: comparing general practice services in a tertiary and primary healthcare settings. *The Nigerian Health Journal*, 17(1):1–14, 2017. URL <https://www.ajol.info/index.php/nhj/article/view/File/154264/143845>. 30.10.2017.
- Paddison, C., Abel, G., Roland, M., Elliott, M., Lyraztopoulos, G., und Campbell, J. Drivers of overall satisfaction with primary care: evidence from the english general practice patient survey. *Health Expectations*, 18(10): 1081 – 1092, 2013.
- Paul, M., Wallace, B., und Dredzde, M. What affects patient (dis)satisfaction? analyzing online doctor ratings with a joint topic-sentiment model. *Workshops at the Twenty- Seventh AAAI Conference on Artificial Intelligence*, page 1 – 6, 2013. URL <http://www.aaai.org/ocs/index.php/WS/AAAIW13/paper/download/7120/6510>. 30.10.2017.
- Pavlič, D., Sever, M., Klemenc-Ketiš, Z., und Švab, I. Process quality indicators in family medicine: results of an international comparison. *BMC Family Practice*, 16:1–11, 2015. URL <https://bmcfampract.biomedcentral.com/track/pdf/10.1186/s12875-015-0386-7?site=bmcfampract.biomedcentral.com>. 30.10.2017.
- Piekenbrock, D. Stichwort: Präferenz. *Gabler Wirtschaftslexikon*, o.J. URL <http://wirtschaftslexikon.gabler.de/Archiv/55419/praefferenz-v5.html>. 05.06.2017.
- Pirk, O. und O., S. Primärdatenerhebung. *Schöffski / Graf v. d. Schulenburg*, page 197 – 242, 2012.
- Porst, R. Wie man die Rücklaufquote bei postalischen Befragungen erhöht. *GESIS- How-to*, 9:1–13, 2001. URL http://www.ssoar.info/ssoar/bitstream/handle/document/20141/ssoar-2001-porst-wie_man_die_rucklaufquote_bei.pdf?sequence=1. 30.10.2017.
- Raab, G., Unger, A., und Unger, F. Methoden der Marketing-Forschung Grundlagen und Praxisbeispiele, 2, 2009. Wiesbaden.
- Rammstedt, B. Reliabilität, Validität, Objektivität. *Wolf / Best*, pages 239 – 258, 2010.
- Ramsay, J., Campbell, J., Schroter, S., Green, J., und Roland, M. The general practice assessment survey (gpas): tests of data quality and measurement properties. *Family Practice*, 17(5):372 – 379, 2000.
- Reimann, S. und Strech, D. The representation of patient experience and satisfaction in physician rating sites. a criteria-based analysis of English- and German-language sites. *BMC Health Services Research*, 12:1–14, 2010. URL <https://bmchealthservres.biomedcentral.com/track/pdf/10.1186/1472-6963-10-332?site=bmchealthservres.biomedcentral.com>. 30.10.2017.
- Richard, E. und Diener, E. Personality and subjective well-being. *Social Indicators Research Series*, 37:75 – 102, 2009.
- Robbins, J., Bertakis, K., L.J., H., Azari, R., Callahan, E., und D.A., C. The influence of physician practice behaviours on patient satisfaction. *Journal of Family Medicine*, 25(1):17 – 20, 1993.
- Rogler, G. Der patient als kunde. *Schweizerische Ärztezeitung*, 90(25):1009 – 1013, 2009.
- Roque, H., Veloso, A., und Ferreira, P. Portuguese version of the EUROPEP questionnaire: contributions to the psychometric validation. *Revista de Saúde Pública*, 50(10):1–7, 2015. URL <http://www.scielo.br/pdf/rsp/v50/0034-8910-rsp-S1518-87872016050006259.pdf>. 30.10.2017.
- Salcher, E. Psychologische Marktforschung, 2, 1995. Berlin.
- Salisbury, C., Wallace, M., und Montgomery, A. Patients' experience and satisfaction in primary care: secondary analysis using multilevel modelling. *BMJ Online*, pages 1–8, 2010. URL <http://www.bmj.com/content/bmj/341/bmj.c5004.full.pdf>. 22.06.2017.
- Schäfer, W. L., Boerma, W. G., Murante, A. M., Sixma, H. J., Schellevis, F. G., und Groenewegen, P. P. Assessing the potential for improvement of primary care in 34 countries: a cross-sectional survey. *Bulletin of the World Health Organization*, 93:161–168, 2015.
- Scharnbacher, K. und G., K. Kundenzufriedenheit: Analyse, Messbarkeit und Zertifizierung, 2003. München.
- Schimanofsky, W. *Prozess- und Kundenorientierung: Analyse – Konzeption – Umsetzung*. Springer-Verlag, 2013.
- Schneider, G., Geiger, I. K., und Scheuring, J. *Prozess- und Qualitätsmanagement: Grundlagen der Prozessgestaltung und Qualitätsverbesserung mit zahlreichen Beispielen, Repetitionsfragen und Antworten*. Compendio Bildungsmedien AG, 2008.
- Schwaiger, M. und Zimmermann, L. Quantitative forschung: Ein überblick. *Manfred Schwaiger und Anton Meyer (Hg.): Theorien und Methoden der Betriebswirtschaft. Handbuch für Wissenschaftler und Studierende*. München:

- Franz Vahlen, pages 420–437, 2011.
- Schäfer, W., Boerma, W., Kringos, D., De Maeseneer, J., Grefß, S., Heinemann, S., Rotar-Pavlic, D., Seghieri, C., Švab, I., Van den Berg, M., Vainieri, M., Westert, G., Willems, S., und Groenewegen, P. QUALICOPC, a multi-country study evaluating quality, costs and equity in primary care. *BMC Family Practice*, 12:1–9, 2011. URL <https://bmcfampract.biomedcentral.com/track/pdf/10.1186/1471-2296-12-115?site=bmcfampract.biomedcentral.com>. 30.10.2017.
- Sebo, P., Herrmann, F., und Haller, D. How do GPs in Switzerland perceive their patients' satisfaction and expectations? An observational study. *BMJ Open*, 5:1–8, 2015. URL <http://bmjopen.bmj.com/content/bmjopen/5/6/e007085.full.pdf>. 30.10.2017.
- Sebo, P., Herrmann, F. R., und Haller, D. M. Is patient satisfaction with organizational aspects of their general practitioner's practice associated with patient and doctor gender? An observational study. *BMC family practice*, 17(1):120 – 127, 2016.
- Sharma, M., Yadav, K., Yadav, N., und Ferdinand, K. C. Zika virus pandemic—analysis of Facebook as a social media health information platform. *American Journal of Infection Control*, 45(3):301–302, 2017.
- Shelton, P. J. *Measuring and improving patient satisfaction*. Jones & Bartlett Learning, Gaithersburg – Maryland, 2000.
- Sitzia, J. und Wood, N. Patient satisfaction: A review of issues and concepts. *Social Science & Medicine*, 45(12):1829 – 1843, 1997.
- Sixma, H., Kerssens, J., Van Campen, C., und Peters, L. Quality of care from the patients' perspective: from theoretical concept to a new measuring instrument. *Health Expectations*, 3(12):82 – 95, 1998.
- Solomon, M. R. und Tuten, T. *Social Media Marketing*. Pearson Higher Ed, New Jersey, 2013.
- Solomon, M. *Konsumentenverhalten*, 2013. München.
- Stauss, B. und Seidel, W. *Beschwerdemanagement: Unzufriedene Kunden als profitable Zielgruppe*. Carl Hanser Verlag GmbH Co KG, 5 edition, 2014.
- Stewart, M. P., Brown, J., Donner, A., McWhinney, I., Oates, J., Weston, W., und Jordan, J. The impact of patient-centered care on outcomes. *The Journal of Family Practice*, 49(9):796 – 804, 2000.
- Swart, E., Ihle, P., Klug, S., und Lampert, T. Gute Praxis Sekundärdatenanalyse (GPS)-Revision nach grundlegender Überarbeitung. *Das Gesundheitswesen*, 70(1):54 – 60, 2008.
- Tachtsoglou, S. und König, J. *Statistik für Erziehungswissenschaftlerinnen und Erziehungswissenschaftler: Konzepte, Beispiele und Anwendungen in SPSS und R*. Springer-Verlag, Wiesbaden, 2017.
- Tahir, M., Nawaz, S., Butt, M., Amin, F., und Tahir-Mahmood, K. In-patient perceptions, needs, expectations and satisfaction within tertiary care settings. *Journal of Pharmaceutical Sciences & Research*, 4(12):2025 – 2031, 2012.
- Tan, S. und Goonawardene, N. Internet health information seeking and the patient-physician relationship: A systematic review. *Journal of Medical Internet Research*, 19(1):1–15, 2017. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5290294/>. 30.10.2017.
- Tarrant, C., Stokes, T., und Baker, R. Factors associated with patients' trust in their general practitioner: a cross-sectional survey. *British Journal of General Practice*, 53(10):798 – 800, 2003.
- Thayaparan, A. und Mahdi, E. The patient satisfaction questionnaire short form (psq-18) as an adaptable, reliable, and validated tool for use in various settings. *Medical Education Online*, 18:1–2, 2013. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3722414/>. 30.10.2017.
- Theis, H. Auf direktem Weg zur Patientenbindung. *Deutsches Ärzteblatt*, 104(13):8 – 12, 2007.
- Theobald, A. *Praxis Online-Marktforschung: Grundlagen-Anwendungsbereiche-Durchführung*. Springer-Verlag, Wiesbaden, 2017.
- Thielsch, M. T. und Weltzin, S. Online-Befragungen in der Praxis. *Praxis der Wirtschaftspsychologie: Themen und Fallbeispiele für Studium und Praxis*, pages 69–85, 2009.
- Thill, K.-D. *Patientenzufriedenheit in der Arztpraxis: die Voraussetzung für eine erfolgreiche unternehmerische Praxisführung*. Deutscher Ärzteverlag, Köln, 2008.
- Thompson, D. und Yarnold, P. Relating patient satisfaction to waiting time perceptions and expectations: The disconfirmation paradigm. *Academic Emergency Medicine*, 2(12):1057 – 1062, 1995.
- Töpfer, A. Kundenzufriedenheit als basis für unternehmererfolg. In *Handbuch Kundenmanagement*, pages 37 – 80. Springer, 2008a.
- Töpfer, A. *Handbuch Kundenmanagement*. Springer-Verlag, Berlin – Heidelberg, 3 edition, 2008b.
- Töpfer, A. Analyse der anforderungen und prozesse wertvoller kunden als basis für die segmentierung und steuerungskriterien. In *Handbuch Kundenmanagement*, pages 191–228. Springer, 2008c.
- Vedsted, P., Sokolowski, I., und Heje, H. Data quality and confirmatory factor analysis of the Danish EUROPEP questionnaire on patient evaluation of general practice. *Scandinavian Journal of Primary Health Care*, 26(3):174 – 180, 2008.
- Vogl, S. Gruppendiskussion. In *Handbuch Methoden der empirischen Sozialforschung*, pages 581–586. Springer, 2014.
- Vukovic, M. H. und Jakovljevic, M. Structural validity and reliability of the healthcare professionals' economic reasoning questionnaire. *Hospital Pharmacology*, 2(3):308 – 316, 2016.
- Vukovic, M. H., Gvozdenovic, B., T. G., Stamatovic Gajic, B., Jakovljevic, M., und B.P. M. Validation of a patient satisfaction questionnaire in primary health care. *Public Health*, 126(8):710 – 718, 2012.
- Ware, J. E. und Snyder, M. K. Dimensions of patient attitudes regarding doctors and medical care services. *Medical Care*, 13(8):669 – 682, 1975.
- Ware, J., Snyder, M., Wright, W., und Davies, A. Defining and measuring patient satisfaction with medical care. *Evaluation and Program Planning*, 6(2):247 – 263, 1983.
- Weis, H. C. und Steinmetz, P. *Marktforschung*, 2008. Ludwigshafen.
- Wengle, H. Qualitäts-standards für die stationäre psychiatrie. *Schweizerische Ärztezeitung*, 81(43):2439 – 2441, 2000.
- Wensing, M., Vedsted, P., Kersnik, J., Peersman, W., Klingenberg, A., Hearnshaw, H., Hjortdahl, P., Paulus, D., Künzi, B., Mendive, J., und Grol, R. Patient satisfaction with availability of general practice: an international comparison. *International Journal of Quality in Health Care*, 14(12):111 – 118, 2002.
- Wildner, M. Patient, Nutzer, Kunde. *Gesundheitswesen*, 73(7):397 – 398, 2011.
- Williams, B. Patient satisfaction: a valid concept. *Social Science and Medicine*, 38(4):509 – 516, 1994.
- Williams, S. und Calnan, M. Key determinants of consumer satisfaction with general practice. *Family Practice*, 18(9):237 – 242, 1991.
- Wong, S., Chau, L., Hogg, W., Teare, G., Miedema, B., Breton, M., Aubrey-Bassler, K., Katz, E. A. / Burge, Boivin, A., Cooke, T., Francoeur, D., und Wodchis, W. An international cross-sectional survey on the Quality and Costs of Primary Care (QUALICO-PC): recruitment and data collection of places delivering primary care across Canada. *BMC Family Practice*, 16:1–10, 2015. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4339081/>. 30.10.2017.
- Wüthrich-Schneider, E. Patientenzufriedenheit – wie verstehen? *Schweizerische Ärztezeitung*, 81(20):1046 – 1048, 2000a.
- Wüthrich-Schneider, E. Patientenzufriedenheit – wie messen? *Schweizerische Ärztezeitung*, 81(21):1116–1119, 2000b.
- Yaphe, J. Demystifying patient satisfaction. *Revista Portuguesa de Medicina Geral e Familiar*, 31(3):162 – 163, 2015.



State-of-the-Art dynamischer Methoden zur multikriteriellen Entscheidungsunterstützung

Sebastian Schär

Georg-August-Universität Göttingen

Abstract

Die Methoden der multikriteriellen Entscheidungsunterstützung (MCDA) bieten die Möglichkeit eine Vielzahl an Kriterien unterschiedlicher Natur im Zuge der Entscheidungsfindung simultan einzubeziehen. Bestimmte Entscheidungen, insbesondere im strategischen Bereich, zeichnen sich zudem durch eine hohe Komplexität aus, da die zugrundeliegenden Annahmen sowie die Auswirkungen der Entscheidung mit Unsicherheiten behaftet sind.

Das Ziel dieser Arbeit war es, durch ein strukturiertes Literaturreview herauszustellen, welche Ansätze zur Erfassung einer solchen dynamischen Entscheidungskomponente es bislang gibt.

Zur Identifikation relevanter Literatur wurden themenrelevante, wissenschaftliche Verlage wie ELSEVIER, sowie die EBS-CO Datenbank genutzt. Auch Dissertationen, Konferenzberichte sowie vorherige Reviewartikel wurden inkludiert. Insgesamt wurden 60 Zeitschriftenartikel aus 31 verschiedenen Zeitschriften, 6 Konferenz-Paper, 11 Buchquellen und eine Dissertation gefunden. Die Literatur wurde anschließend nach dem zugrundeliegenden Verständnis der dynamischen Komponente, sowie deren methodischer Erfassung klassifiziert. Hierbei offenbarten sich drei Gruppen von Ansätzen Dynamik in die MCDA zu integrieren: (1) Szenario-basierte Ansätze, (2) Eine Kombination von MCDA mit Lebenszyklusmodellen (LCA), sowie (3) die direkte Einbeziehung von Dynamik in der Problemformulierung über mehrere Datensätze (DMCDA).

Ein kritischer Vergleich dieser zeigt eine fortgeschrittene Entwicklung mit vielen Anwendungsbeispielen im Forschungsstrang der Szenario-basierten Ansätze. Eine Kombination von MCDA mit LCA kommt vor allem in Nachhaltigkeitsfragen und bei der Beurteilung von Energietechnologien zum Einsatz. Das Gebiet der DMCDA-Ansätze erweist sich als vergleichsweise jüngerer Forschungsstrang mit Ansatzpunkten für zukünftige Forschungsvorhaben.

Keywords: Multikriterielle Entscheidungsunterstützung, DMCDA, uncertainty, dynamic decision making, MADM

1. Einleitung

Die multikriterielle Entscheidungsunterstützung (engl. Multi-Criteria Decision Analysis, MCDA) erfreut sich in den letzten Dekaden wachsender Beliebtheit. Im Gegensatz zu traditionellen Modellen der Entscheidungsunterstützung, welche häufig nur ein entscheidungsrelevantes Kriterium berücksichtigen, bieten MCDA-Methoden die Möglichkeit eine Vielzahl an Kriterien aus unterschiedlichen Bereichen zu berücksichtigen. Neben quantitativen Kriterien können somit auch gleichzeitig qualitative Aspekte einbezogen werden. Insbesondere im Umgang mit komplexen Entscheidungsproblemen können Methoden der multikriteriellen Entscheidungsunterstützung durch eine strukturierte Problemaufbereitung eine Grundlage für fundierte Entscheidungen liefern (Zimmermann und Gutsche (1991)).

Für bestimmte Entscheidungen ist es zudem hilfreich,

ebenfalls eine zukunftsorientierte Komponente in die Entscheidungsfindung einzubeziehen. Vor allem strategische Planungsentscheidungen unterliegen einer gewissen Dynamik, da die Rahmenbedingungen und Annahmen, unter denen die Entscheidung getroffen wird, sich im Zeitverlauf entscheidend ändern können. Auch die Auswirkungen der getroffenen Entscheidung sind oftmals nicht genau vorhersehbar und somit mit einem gewissen Risiko behaftet (Stewart et al. (2013)). Diese zeitlichen Aspekte gilt es bei dynamischen MCDA-Methoden zu berücksichtigen.

Da die Zahl der Veröffentlichungen zu Methoden der multikriteriellen Entscheidungsunterstützung in den vergangenen Jahren stark angestiegen ist und die Arbeiten zunehmend differenzierter werden, bietet sich das Verfassen eines Reviews an, um die Ergebnisse oder eventuelle Inkonsistenzen und Widersprüche dynamischer MCDA-Methoden aufzu-

zeigen, und so eine Basis für zukünftige Arbeiten und Praxisprojekte zu schaffen (Fettke (2006); Ehr Gott et al. (2010)). Gegenstand dieser Arbeit ist es, mithilfe einer strukturierten Literaturrecherche herauszuarbeiten, welche Ansätze es bislang in diesem Forschungsfeld gibt, um zeitliche Aspekte in MCDA-Methoden zu integrieren.

Die Arbeit wird in einen Grundlagenteil, dem Kapitel zwei, und in einen Hauptteil, den Kapiteln drei bis fünf unterteilt. Kapitel sechs schließt die Arbeit ab.

Kapitel zwei soll sich mit den theoretischen Grundlagen der multikriteriellen Entscheidungsunterstützung befassen. Hierzu werden zunächst Bestandteile von Entscheidungsproblemen sowie der Prozess der Entscheidungsfindung mitsamt den gängigen Methoden der multikriteriellen Entscheidungsunterstützung vorgestellt. Auch die mathematische Darstellung soll grundlegend erläutert werden, um eine inhaltliche Auseinandersetzung der recherchierten Methoden im späteren Verlauf der Arbeit zu ermöglichen. Ferner ist das Konzept der Dynamik zu definieren.

Das anschließende Kapitel drei soll das methodische Vorgehen der Arbeit verdeutlichen. Zum Erreichen einer möglichst repräsentativen Stichprobe gilt es, geeignete Suchbegriffe und -kriterien festzulegen, die eine möglichst gezielte Suche nach dynamischen Methoden ermöglichen. Ebenso soll durch eine Charakterisierung dieses Reviews nach dem Schema von Fettke (2006) ein konsistentes und reproduzierbares Vorgehen gewährleistet werden.

In Kapitel vier erfolgt schließlich die Aufbereitung der Rechercheergebnisse. Hier gilt es herauszuarbeiten, wie die dynamische Komponente in den ausgewählten Arbeiten definiert wird und auf welche Weise sie methodische Berücksichtigung findet.

In Kapitel fünf erfolgt eine kritische Diskussion der Rechercheergebnisse. Hierzu werden einerseits Gemeinsamkeiten der ausgesuchten Arbeiten herausgestellt und integriert, als auch Auffälligkeiten analysiert und Lücken in der Forschung aufgezeigt.

Die aus den vorherigen Kapiteln gewonnenen Erkenntnisse werden im abschließenden Kapitel sechs generalisiert, diskutiert und konkreten Handlungsempfehlungen für weitere Arbeiten zu dynamischen Methoden der multikriteriellen Entscheidungsunterstützung vorangestellt.

2. Theoretische Grundlagen multikriterieller Entscheidungsunterstützungssysteme

In diesem Kapitel werden die begrifflichen Grundlagen für die weitere Arbeit geschaffen. Hierzu wird in Kapitel 2.1 der Begriff der multikriteriellen Entscheidungsunterstützungssysteme definiert und erläutert, worin Funktion, Nutzen und Ziele dieser bestehen. Anschließend wird die Grundstruktur von Entscheidungsproblemen und der Prozess der Entscheidungsfindung im Ablauf erklärt. Abgeschlossen wird das Kapitel 2.1 durch einen schematischen Überblick gängiger Methoden der multikriteriellen Entscheidungsunterstützung. Das Kapitel 2.2 soll für eine spätere Verwendung

kurz und bündig in die mathematische Notation multikriterieller Entscheidungsprobleme einführen. Schließlich soll eine Definition für das Konzept der Dynamik in Zusammenhang mit multikriteriellen Entscheidungsunterstützungssystemen gefunden werden. Um der Aufgabe eines Reviews gerecht zu werden, wird eine Abwägung über den Umfang der gewählten Definition an dieser Stelle entscheidend sein. Zum einen ist der Begriff eindeutig und klar zu definieren, zum anderen gilt es zu verhindern, dass eventuelle Schlüsselarbeiten aufgrund einer abweichenden Definition des dynamischen Aspektes aus der Betrachtung ausgeschlossen werden.

2.1. Entscheidungstheoretische Grundlagen

Bereits alltägliche Situationen, wie die Wahl des Abendessens, stellen ein multikriterielles Entscheidungsproblem dar. Preis, Geschmack, Ort des Einkaufs oder der anfällige Zeitaufwand sind nur einige der vielen Kriterien, welche bei dieser Entscheidung eine Rolle spielen (Yu und Chen (2010)). Ebenso gilt es bei ökonomischen Entscheidungen nicht nur ein einziges kosten- oder gewinnorientiertes Kriterium zu berücksichtigen, sondern vielmehr eine Reihe an relevanten Zielen, welche möglicherweise sogar in Konkurrenz zueinanderstehen und aufgrund unterschiedlicher Maßstäbe nicht problemlos vergleichbar sind (Zimmermann und Gutsche (1991)). Nach Zimmermann und Gutsche sind folgende grundlegenden Charakteristika kennzeichnend für multikriterielle Entscheidungsprobleme:

- **Mehrere Ziele:** Der Entscheidungsträger entscheidet hierbei selbst über die Relevanz eines Ziels für das Entscheidungsproblem.
- **Zielkonflikt:** Eine Verbesserung hinsichtlich eines Ziels verschlechtert das Ergebnis hinsichtlich eines anderen Ziels. Es wird ein sinnvolles Kriterium zur Beurteilung der Alternativen benötigt.
- **Unvergleichbare Einheiten:** Ziele werden mit unvergleichbaren Maßstäben bzw. in nicht vergleichbaren Einheiten gemessen
- **Auswahl einer Lösung:** Die Handlungsalternative, welche der Entscheidungsträger im Hinblick auf alle Ziele am meisten bevorzugt, wird ausgewählt. Ist die Menge an Handlungsalternativen endlich, so wird die beste Alternative ausgewählt. Bei nicht explizit festgelegter Anzahl an Handlungsalternativen wird die beste Handlungsalternative durch Berechnung ausfindig gemacht.

2.1.1. Multikriterielle Entscheidungsunterstützung

Belton und Stewart (2002) bezeichnen MCDA als Überbegriff für eine Vielzahl von formalen Herangehensweisen eines Entscheidungsträgers oder einer Gruppe von Entscheidungsträgern an Probleme mit mehr als einem relevanten Zielkriterium. Dabei hat MCDA ausdrücklich nicht den Anspruch

eine optimale¹ Lösung zu generieren, welche den oder die Entscheidungsträger von der eigenen Verantwortung befreit. Vielmehr ist die unterstützende Wirkung des Prozesses hervorzuheben, welcher den Entscheidungsträgern dabei hilft, verfügbare Informationen zu organisieren und die subjektive Komponente eines Entscheidungsprozesses gezielt zu formulieren, um ein besseres Verständnis für das Problem zu erlangen (Belton und Stewart (2002); Wilkens (2012)). Vor allem bei schwierigen Gruppenentscheidungen kann dieses Vorgehen das Finden einer gemeinsamen Lösung erleichtern, die Akzeptanz dieser steigern und eventuell sogar neue Handlungsalternativen aufzeigen (Phillips (1990)).

Innerhalb der MCDA wird zwischen den Methoden des MODM (engl. Multi-Objective Decision Making) und des MADM (engl. Multi-Attribute Decision Making) unterschieden. MODM-Methoden bestimmen mithilfe mathematischer Verfahren durch die Aufstellung von Restriktionen eine optimale Lösung aus einem stetigen Lösungsraum. Bei MADM-Methoden hingegen ist die diskrete Anzahl an Alternativen bereits vor Anwendung der Methode bekannt und das Ziel der Methode liegt in einem Vergleich der verschiedenen Alternativen im Hinblick auf definierte Kriterien (Zimmermann und Gutsche (1991)). Weiterhin wird innerhalb der MADM-Methoden zwischen zwei Ansätzen unterschieden: Den Ansätzen der amerikanischen Schule und denen der europäischen. Für letztere findet sich in der Literatur teilweise die Bezeichnung Multi-Criteria Decision Aid, abgekürzt mit MCDA (Omann (2004); Wilkens (2012)). Es sei jedoch ausdrücklich erwähnt, dass im Rahmen dieser Arbeit die Abkürzung MCDA stets den übergeordneten Begriff für sämtliche oben genannten Methoden der multikriteriellen Entscheidungsunterstützung bezeichnet. Vergleichbare Auffassungen dieses Begriffs finden sich etwa bei Belton und Stewart (2002) und Polatidis et al. (2006). An anderer Stelle in der Literatur wird diese Trennung weniger explizit vorgenommen und die Begriffe werden oftmals synonym verwendet (Figueira et al.). Dieses Review beschäftigt sich vorrangig mit den Möglichkeiten zur Implementation der Dynamik innerhalb der Methoden des MADM. Den Methoden des MODM kommt demnach nur eine ergänzende Rolle zu.

Bestandteile eines multikriteriellen Entscheidungsproblems

Um die Bestandteile der MCDA systematisch und konsistent zu erfassen, findet sich in der Literatur die sogenannte CAUSE Checkliste (engl. Criteria, Alternatives, Uncertainty, Stakeholders, Environment). Mithilfe dieser Merkregel lassen sich die folgenden fünf wesentlichen Strukturelemente der MCDA kurz und bündig erklären (Belton und Stewart (2002); Diakoulaki und Grafakos (2004); Ehr Gott et al. (2010)):

Anhand von den Entscheidungsträgern festgelegter Kri-

terien (engl. Criteria) werden die zur Verfügung stehenden Alternativen (engl. Alternatives) verglichen. Eine Alternative bezeichnet in diesem Zusammenhang eine Handlungsoption oder Strategie, welche vor Anwendung der MCDA systematisch erarbeitet und definiert werden sollte (Zimmermann und Gutsche (1991)). Aufgrund der iterativen Natur des MCDA-Prozesses können sich nachträglich jedoch auch noch weitere Alternativen entwickeln und von den Entscheidungsträgern hinzugefügt werden. Die zum Vergleich herangezogenen Kriterien sollten die Ziele und Wertvorstellungen der involvierten Entscheidungsträger widerspiegeln, möglichst eindeutig definiert und mit quantitativen oder qualitativen Skalen messbar sein. Da im Rahmen dieser Arbeit vor allem komplexe Entscheidungsprobleme mit strategischem Charakter betrachtet werden, bezeichnet ein Entscheidungsträger im Zusammenhang dieser Arbeit nur diejenigen Personen, welche den MCDA-Prozess aktiv gestalten und letztendlich auch durch Auswahl einer Alternative Verantwortung für eine Entscheidung tragen. Es kann lediglich ein einzelner, aber auch mehrere Entscheidungsträger im Entscheidungsprozess involviert sein. Der Gruppe der Stakeholder hingegen gehören neben den Entscheidungsträgern selbst zudem jegliche Personen an, die ein Interesse am Entscheidungsproblem besitzen beziehungsweise von den Auswirkungen einer Entscheidung betroffen sind.

Unsicherheit (engl. Uncertainty) ist ein weiterer wesentlicher Bestandteil (dynamischer) multikriterieller Entscheidungsprobleme. Hierbei ist zwischen interner und externer Unsicherheit zu unterscheiden (Wilkens (2012)). Interne Unsicherheit tritt in Zusammenhang mit unpräzise formulierten Daten, wie nicht skalierten linguistischen Äußerungen, auf (Stewart und Durbach (2016)). Diese Form von Unsicherheit kann mithilfe von sogenannten „fuzzy sets“² methodisch verarbeitet werden (Zimmermann und Gutsche (1991)). Wesentlicher für diese Arbeit ist jedoch die externe Unsicherheit. Werden im Rahmen der Entscheidungsfindung Annahmen über zukünftige Entwicklungen nicht unter vollkommener Sicherheit getroffen, ist damit stets ein Risiko verbunden. Es ist daher eine wichtige Aufgabe der Entscheidungsträger, diese Unsicherheit möglichst genau zu erfassen und sich dem damit verbundenen Risiko bewusst zu sein (Stewart und Durbach (2016)).

Jede Entscheidungsfindung findet zudem unter bestimmten variablen zeitlichen und örtlichen Umweltfaktoren (engl. Environment) statt, welche das Entscheidungsproblem zusätzlich beeinflussen. Dies können unter anderem staatliche Vorgaben, sozioökonomische oder auch kulturelle Aspekte sein (Diakoulaki und Grafakos (2004)).

Ablauf des Entscheidungsfindungsprozesses

Der Entscheidungsfindungsprozess ist in mehrere Stufen

¹Es wird ferner ausdrücklich erwähnt, dass im Kontext multikriterieller Entscheidungen das Konzept der Optimalität keine Anwendung findet und somit generell keine eindeutig optimalen Entscheidungen in diesem Rahmen existieren (Belton und Stewart (2002)).

²unscharfe Mengen (engl. fuzzy sets) versuchen die fehlende Präzision der verbalen Ausdrucksweise zu erfassen. Unschärf bewertetete Alternativen, wie etwa „sehr teuer“ oder „eher teuer“, werden dabei durch Zugehörigkeitsfunktionen den Mengen der „teuren“ Alternativen mit jeweils einem unterschiedlichen Grad der Zugehörigkeit zugeteilt (Great Britain, Department for Communities and Local Government (2009)).

unterteilt und folgt häufig folgendem generellen Schema (Belton und Stewart (2002); Wilkens (2012)):

1. Beschreibung des Entscheidungskontextes: Bestimmung der Entscheidungsträger, Stakeholder und Ziel des Entscheidungsprozesses
2. Festlegung der zu bewertenden Alternativen
3. Festlegung der Ziele und Kriterien sowie Werte der Alternativen
4. Beschreibung der erwarteten Zufriedenstellung jeder Alternative durch Bewertung
5. Gewichtung der Kriterien entsprechend ihrer relativen Wichtigkeit für die Zielerreichung
6. Aggregation der Gewichtungen und Bewertungen für jede Alternative, um einen Gesamtwert zu erhalten
7. Auswertung der Ergebnisse
8. Sensitivitätsanalysen

Zwar erfolgt ein Entscheidungsfindungsprozess schrittweise nach diesem grundlegenden Schema, jedoch ist er nicht als starr und geradlinig anzusehen (Diakoulaki und Grafakos (2004)). Bis zur endgültigen Entscheidungsfindung können einzelne Schritte oder Schrittfolgen auch wiederholt durchlaufen werden, was dem Vorgehen einen eher iterativen Charakter zukommen lässt (Belton und Stewart (2002)).

2.2. Mathematische Formalisierung multikriterieller Entscheidungsprobleme

In diesem Abschnitt wird ein kurzer Überblick über die mathematische Modellierung von MCDA Problemen, basierend auf den Arbeiten von Chen et al. (2010), Campanella und Ribeiro (2011) sowie Jassbi et al. (2014) gegeben. Hierdurch wird die Grundlage für eine spätere Auseinandersetzung mit der Modellerweiterung um eine dynamische Komponente geschaffen.

Die insgesamt m Alternativen werden mithilfe der Variable a_i , wobei $i = 1, \dots, m$, abgebildet. Die n verschiedenen Kriterien mithilfe der Variable c_j , wobei $j = 1, \dots, n$. Eine normalisierte Variable $x_{ij} \in [0, 1]$ repräsentiert das Ausmaß der Zufriedenstellung einer Alternative a_i bezüglich eines Kriteriums c_j , oftmals auch als Konsequenz bezeichnet. Ein Wert von 1 korrespondiert hier mit einer maximalen Zufriedenstellung. Folglich ergibt sich folgende $m \times n$ -Matrix für das multikriterielle Entscheidungsproblem:

$$\begin{matrix} & c_1 & \dots & c_n \\ \begin{matrix} a_1 \\ \vdots \\ a_m \end{matrix} & \begin{bmatrix} x_{11} & \dots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \dots & x_{mn} \end{bmatrix} & = & \begin{bmatrix} X_1 \\ \vdots \\ X_m \end{bmatrix} \end{matrix} \tag{1}$$

Präferenzen bezüglich Kriterien werden als Gewichtungen (engl. weight) bezeichnet. Die relative Wichtigkeit eines Kriteriums c_j aus Sicht der Entscheidungsträger wird durch einen Gewichtungsfaktor w_j mit $j = 1, \dots, n$ dargestellt. Aus allen Gewichtungsfaktoren ergibt sich letztlich der Gewichtsvektor $W \in [0, 1]^n$, mit $\sum_{j=1}^n w_j = 1$.

Präferenzen bezüglich einer Konsequenz x_{ij} werden als Werte (engl. value) bezeichnet und sind abhängig von den individuellen Zielen des Entscheidungsträgers. Um eine Evaluation zu ermöglichen ist es notwendig, alle Konsequenzen x_{ij} anhand dieser Werte zu konvertieren:

$$v_{ij} = f(x_{ij}) \tag{2}$$

Aus (2) ergibt sich für jede Alternative a_i der Wertevektor $v(a_i) = (v_{i1}, v_{i2}, \dots, v_{in})$. In einem abschließenden Schritt wird nun mithilfe einer Aggregationsfunktion ein globaler Ergebnisvektor $\chi \in [0, 1]^n$ erzeugt. Dieser gibt für jede Alternative einen Wert X_i entsprechend der Vorteilhaftigkeit zwischen 0 und 1 aus:

$$\chi = X(a_i) = F(v(a_i), w) \tag{3}$$

Anhand von (3) lässt sich erkennen, dass die Wahl einer geeigneten Aggregationsfunktion von entscheidender Bedeutung für das Ergebnis ist. Diese Thematik soll an späterer Stelle weiter ausgeführt werden.

2.3. Definition des dynamischen Aspektes in multikriteriellen Entscheidungsunterstützungssystemen

Aus der in Kapitel 2.1 bereits angesprochenen Unsicherheit im Zusammenhang mit Entscheidungsproblemen ergibt sich die Notwendigkeit, diesen Bestandteil methodisch zu verarbeiten. Bereits für Keynes zeichnete sich Unsicherheit dadurch aus, dass ihr nicht sinnvoll mit einer Bestimmung von Wahrscheinlichkeiten begegnet werden kann:

“[Under uncertainty] there is no scientific basis on which to form any calculable probability whatever. We simply do not know. Nevertheless, the necessity for action and for decision compels us as practical men to do our best to overlook this awkward fact [...]” (Keynes (1937); nach Montibeller et al. (2006))

Eine ähnliche Definition von Unsicherheit findet sich auch bei Crousillat et al. (1993), hier ausdrücklich im Kontext von Entscheidungsproblemen mit mehreren Alternativen und konkurrierenden Zielen. Jedoch lässt sich mit dem Begriff der Unsicherheit allein der dynamische Aspekt nicht ausreichend präzise definieren. Für eine methodische Berücksichtigung besteht also weiterer Charakterisierungsbedarf. In Gerking (1987) nennt der Autor verschiedene Quellen der Unsicherheit und kombiniert das Konzept der Unsicherheit mit dem der Dynamik in einer sogenannten „dynamischen Unsicherheit“. Diese definiert er als eine im Zeitverlauf sich durch äußere Entwicklungen reduzierende Unsicherheit. Der Unsicherheitsaspekt wird also um eine zeitlich abhängige Komponente erweitert.

Dieses Verständnis der „dynamischen Unsicherheit“ trifft man in der Literatur häufig im Zusammenhang mit Nachhaltigkeit und nachhaltigen Entscheidungen an. Hier kann Unsicherheit zum einen seitens der Entscheidungsträger bezüglich der eigenen Überzeugungen bestehen, aber auch im Hinblick auf die Auswirkungen von Entscheidungen, die häufig nicht mithilfe von Wahrscheinlichkeiten quantifiziert werden können (Lienert et al. (2015); Reichert et al. (2015)). Ein weiterer Unsicherheitsfaktor ist zudem die zukünftige Entwicklung externer Einflussfaktoren (Ribeiro et al. (2013); Scholten et al. (2014)). Montibeller et al. (2006) sehen eine Notwendigkeit für die Entscheidungsträger, sich den wichtigsten Quellen der Unsicherheit bewusst zu werden und ihnen methodisch mittels Ausgestaltung von möglichen Szenarien zu begegnen. Ein ähnliches Verständnis der Dynamik findet sich bei Stewart et al. (2013) und Frini (2014). Sie begreifen die Dynamik als alle verschiedenen Möglichkeiten, wie sich unsichere Zusammenhänge entwickeln können. Die Aufgabe der MCDA ist es, diese bereits in der Gegenwart in den Entscheidungsprozess einfließen zu lassen und so für nachhaltige Entscheidungen zu sorgen.

Weitere Unsicherheiten bestehen bezüglich des technologischen Wandels (Myllyviita et al. (2012)). Die meisten Technologien unterliegen durch komplexe Entwicklungen im Zeitverlauf einem stetigen Wandel. Während neue Technologien in ihrer Leistung und unter Nachhaltigkeitsaspekten den etablierten zwar zunächst unterlegen sein mögen und ihre Einführung zunächst mit Risiken und Unsicherheiten verbunden ist, können sie trotzdem das Potential für eine größere Leistungsfähigkeit besitzen. Dieser Komponente der Dynamik versuchen Oberschmidt et al. (2010) methodisch mithilfe von MCDA und Lebenszyklusmodellen zu begegnen. Eine wichtige Eigenschaft dieser MCDA-Modelle besteht hier in einer Lernfähigkeit der Methode für zukünftige Entscheidungen (Cinelli et al. (2013)). Halog und Manik (2011) definieren den dynamischen Aspekt in diesem Zusammenhang eng angelehnt an die Methode der ‚System Dynamics‘³.

Eine erste ausdrückliche Verwendung des Begriffs der dynamischen multikriteriellen Entscheidungsunterstützung (engl. Dynamic Multi-Criteria Decision Making) findet sich in der Arbeit von Kornbluth (1992). Er erweitert die zum damaligen Zeitpunkt vorliegenden Ansätze durch das Einbeziehen einer möglichen Präferenzänderung der Entscheidungsträger im Zeitverlauf und die daraus resultierende Auswirkung auf die gegenwärtige Entscheidungsfindung. So sind laut Kornbluth weder Kriteriengewichtungen noch Wertvorstellungen als statisch anzusehen, sondern unterliegen vielmehr durch das Vorliegen von Unsicherheit einem stetigen Wandel. Die Dynamik liegt also in der permanenten Interaktion von Kriterien und Alternativen (Corner et al. (2001)) im Rahmen eines explorativen Prozesses der Entscheidungsträger,

an dessen Ende erst die finale Entscheidung steht. Zudem beeinflussen vergangene Denkprozesse unterbewusst spätere Entscheidungen (Campanella und Ribeiro (2011)). Chen et al. (2010) begreifen DMCDa (engl. Dynamic Multi-Criteria Decision Analysis) als eine Erweiterung der herkömmlichen MCDA durch die Verwendung mehrerer Datensätze zu Alternativen und Kriterien über einen gewissen Zeitraum und deren methodische Einbeziehung in den Entscheidungsprozess. So sind die involvierten Parameter wie Kriterien, Alternativen, Entscheidungen, Präferenzen oder aber die Problemstruktur nicht länger als fix und stetig anzunehmen, sondern können situationsbedingt oder im Zeitverlauf Änderungen unterliegen (Yu und Chen (2010); Jassbi et al. (2014)).

Angelehnt an diese Definition wird auch die Arbeitsdefinition des dynamischen Aspektes in diesem Review gewählt: Neben einer, von den Entscheidungsträgern ausgehenden, zeitbezogenen Unsicherheit inkludiert der Begriff ebenso exogene Dynamiken im Zusammenhang mit dem technologischen Wandel und im Kontext nachhaltiger Entwicklung. Diese Dynamik spiegelt sich letztendlich methodisch durch sich im Zeitverlauf ändernde Parameter und der Verwendung historischer Daten für zukünftige, zu treffende Entscheidungen wider.

3. Methodik der Literatursuche

“An effective review creates a firm foundation for advancing knowledge. It facilitates theory development, closes areas where a plethora of research exists, and uncovers areas where research is needed.” (Webster und Watson (2002))

Um dieser Aufgabe möglichst angemessen nachzukommen, wird in diesem Kapitel eine methodische Vorgehensweise für den folgenden Verlauf der Arbeit entwickelt. Ein etabliertes Schema für die Erstellung eines Reviewbeitrages präsentiert Fettke (2006) in seiner Arbeit nach Vorbild der fünf Phasen der Reviewforschung von Cooper und Hedges (1994, S. 9–13). Dieses Vorgehen soll auch bei der Erstellung dieser Arbeit als methodischer Handlungsrahmen dienen und weitestgehend Reproduzierbarkeit und ein konsistentes Vorgehen gewährleisten:

Nachdem die erste Phase aus Abbildung 1 mit dem Kapitel eins und zwei dieser Arbeit bereits abgeschlossen wurde, steht in diesem Kapitel nun die Recherche relevanter Literatur an. Um den Suchprozess möglichst zielorientiert gestalten zu können, bietet es sich an dieser Stelle an, zunächst die spezifischen Anforderungen an dieses Review herauszuarbeiten. Neben dem Umfang der zu untersuchenden Stichprobe gilt es unter anderem zu klären, an welchen Personenkreis das Review adressiert ist, da dies den Fokus der Untersuchung entscheidend beeinflusst. Des Weiteren soll auch das zu verfolgende Ziel der Arbeit konkretisiert werden. Dies kann etwa in einer Integration der vorliegenden Literatur, aber auch in einer kritischen Diskussion oder einer Herausarbeitung zentraler Aspekte bestehen. Deshalb wird in Kapitel

³ ‚System Dynamics‘ ist eine von Jay Forrester am MIT entwickelte wissenschaftliche Methode zur Untersuchung und Modellierung der Beziehungen zwischen und innerhalb von Subsystemen. Ziel ist es, die Auswirkung dieser Beziehungen auf das Verhalten des Gesamtsystems herauszufinden (Halog und Manik (2011)).



Abbildung 1: Fünf Phasen der Reviewforschung; Quelle: [eigene Darstellung nach Fettke (2006)]

3.1 eine Charakterisierung dieses Reviews nach dem Schema von Fettke (2006) vorgenommen.

Anschließend werden darauf aufbauend in Kapitel 3.2 geeignete Suchbegriffe und Vorgehensweisen bei der Literaturrecherche vorgestellt. Es sollen gezielt Arbeiten gesucht werden, welche den dynamischen Aspekt von MCDA-Methoden berücksichtigen. Fraglich ist, ob Autoren diesen ebenfalls in konsistenter Weise als „dynamisch“ kennzeichnen oder ob unterschiedliche Bezeichnungen in der bisherigen Forschung verwendet werden. Die verwendeten Suchbegriffe sollten das Konzept der Dynamik, im Rahmen der in Kapitel 2.3 festgelegten Definition, also möglichst ganzheitlich erfassen und somit den methodischen Rahmen für die anschließende Literatursuche vervollständigen.

3.1. Charakteristik des Reviews

Der Begriff eines Reviews ist in der Literatur verhältnismäßig weit gefasst und wird je nach Autor unterschiedlich definiert (Gardfield (1987), S. 114ff.). Aus diesem Grund präsentiert Fettke in seiner Arbeit folgendes Schema zur Differenzierung verschiedener Typen von Reviews (Fettke (2006) nach: Cooper (1988); Light und Pillemer (1984); Mantel (1973); Virgo (1971)):

Die Einteilung in das Charakterisierungsschema von Fettke, aufgezeigt in Abbildung 2, stellt die Ansprüche an dieses Review heraus und soll ein konsistentes Vorgehen gewährleisten. Ziel ist, nicht zu jedem charakteristischen Kriterium strikt eine Kategorie zu wählen, sondern durch schematisches Vorgehen anschließend eine klar definierte Anforderung an das Review zu erhalten. Ein Review kann demnach einzelne Aspekte isoliert betrachten, es sind jedoch auch kombinierte Betrachtungen denkbar.

Hinsichtlich der Kategorie Typ bietet sich ein natürlichsprachliches Review an. Der Anspruch dieser Arbeit besteht nicht in einer rein quantitativen Betrachtung des derzeitigen State-of-the-Art dynamischer MCDA-Methoden, sondern soll vor allem qualitativ den derzeitigen Forschungsstand beleuchten. Mathematisch-statistische Instrumente nehmen dabei höchstens eine unterstützende Rolle ein.

Resultierend wird der Fokus auf die Theorie im weiteren Sinne, also Bezugsrahmen und Konzepte zur Einbringung der dynamischen Komponente als theorieähnliche Artefakte (Fettke (2006), S. 259), sowie die darauf aufbauenden Methoden, gelegt.

Das Ziel dieses Reviews wurde bereits expliziert, es bedarf jedoch noch einer inhaltlichen Kategorisierung. Im Vordergrund steht hierbei herauszufinden, welche zentralen Ansätze es bisher in der Forschung gibt, um zeitliche Aspekte

in MCDA-Methoden einfließen zu lassen. Neben dieser Herausarbeitung von zentralen Aspekten bietet sich zudem eine kritische Bewertung des derzeitigen Forschungsstandes an, um gezielt Anwendungsmöglichkeiten und Lücken aufzuzeigen. Eine Integration erscheint aufgrund der teilweise stark unterschiedlichen Konzepte zur zeitlichen Berücksichtigung in MCDA-Methoden als nicht besonders sinnvoll und für den Rahmen dieser Arbeit zu umfangreich.

Dieses Review wird aus einer neutralen Perspektive verfasst. Das heißt, es wird bei der Durchsicht der betrachteten Literatur keine theoretische Position bezogen. Auch unterliegt der Zielstellung dieser Arbeit keine Annahme, welche es zu verifizieren oder falsifizieren gilt. Vielmehr wird der thematische Gegenstand dieser Arbeit methodisch von Grund auf neu erarbeitet.

Um diese methodische Erarbeitung möglichst transparent und nachvollziehbar zu gestalten, ist es zudem unabdingbar, die Literatúrauswahl explizit zu erläutern und zu begründen. Dabei sollte die ausgewählte Literatur möglichst repräsentativ für die vorgestellten Ansätze zur Integration der dynamischen Komponente sein. Ziel der Literatúrauswahl ist die Recherche mindestens eines repräsentativen Forschungsstrangs zu jedem der präsentierten Ansätze und so den gegenwärtigen Stand der Forschung in diesem Gebiet herauszustellen.

Die Struktur dieser Arbeit wird demnach unter methodischen Gesichtspunkten gestaltet. Es werden jeweils Arbeiten mit ähnlichen methodischen Ansätzen zur Integration der zeitlichen Komponente vergleichend untersucht und kritisiert.

Die Arbeit ist an eine Zielgruppe, bestehend aus Anwendern von MCDA-Methoden und somit Praktikern als auch Forscher im Gebiet der Operations Research⁴, gerichtet.

Anknüpfungspunkte für zukünftige Forschung sollen durch einen kritischen Vergleich in Kapitel 5 der Arbeit expliziert werden.

3.2. Quellenauswahl und Suchmethodik

In diesem Kapitel stehen Phase zwei und drei des Reviewprozesses nach Fettke im Mittelpunkt. Es gilt, zur Klärung der Fragestellung geeignete Literatur zu finden und die dabei verwendete Methodik darzulegen sowie die gefundene Literatur zu verarbeiten.

Zum Erreichen einer möglichst repräsentativen Stichprobe bedarf es geeigneter Suchbegriffe und -kriterien, die eine

⁴Eine genauere Erläuterung des Begriffs der Operations Research findet sich in Kallrath (2013, S. 3ff.).

Charakteristik		Kategorie			
1. Typ		natürlichsprachlich		statistisch	
2. Fokus		Forschungsergebnis	Forschungsmethode	Theorie	Erfahrung
3. Ziel	Nennung	nicht expliziert		expliziert	
	Inhalt	Integration	Kritik	zentrale Themen	
4. Perspektive		neutral		Position	
5. Literatur	Auswahl	nicht expliziert		expliziert	
	Umfang	Schlüsselarbeiten	repräsentativ	selektiv	vollständig
6. Struktur		historisch	thematisch	methodisch	
7. Zielgruppe		Allgemeine Öffentlichkeit	Praktiker	Forscher im Allgemeinen	spezialisierte Forscher
8. Zukünftige Forschung		nicht expliziert		expliziert	

Abbildung 2: Schematische Einteilung des Reviews; Quelle: [eigene Darstellung nach Fettke (2006)]

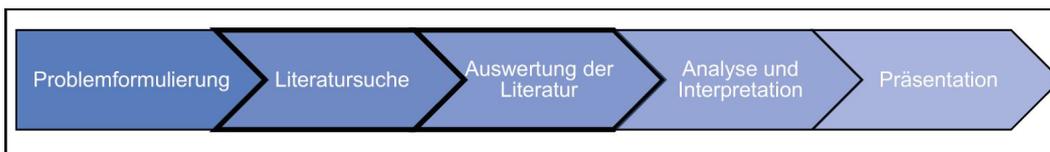


Abbildung 3: Aktuelle Phase des Reviewprozesses; Quelle: [eigene Darstellung nach Fettke (2006)]

möglichst gezielte Suche nach dynamischen Methoden ermöglichen. Hierzu merken Cooper und Hedges (1994) an, dass sich selbst bei der Nutzung der bekanntesten großen Datenbanken noch Erkenntnislücken durch unveröffentlichte Arbeiten ergeben können. Diese ließen sich nur durch Kontaktierung der auf das Themengebiet spezialisierten Forscher inkludieren, was einer objektiven und erstrebenswerten Suchmethodik jedoch nicht gerecht werden würde. Anstelle dem Anspruch der Vollständigkeit nachzukommen, sollte der Fokus für den Reviewsynthetiker somit eher auf der Wahl möglichst komplementärer und sorgfältig ausgewählter Erkenntnisquellen liegen. Auch für dieses Review dient dies als Leitlinie der Literatursuche.

Als Erkenntnisquellen kommen somit die Suchmasken der themenrelevanten wissenschaftlichen Verlage wie ELSEVIER, sowie die Datenbank der Ebsco Business Source Premier (EBSCO) in Frage. Lehrbüchern fällt aufgrund der fehlenden Aktualität nur eine ergänzende Rolle zu. Aufgrund der hohen Themenspezifität und der Tatsache, dass es sich bei dem Gebiet der dynamischen MCDA-Methoden um ein relativ junges Forschungsfeld handelt, soll das Ranking nach dem Verband der Hochschullehrer für Betriebswirtschaft e.V. (vhbonline.org) der in Betracht kommenden Journals zwar betrachtet werden, aber zunächst nur ein zweiträn-

giges Klassifizierungskriterium sein. Auch Dissertationen, veröffentlichte Konferenzberichte sowie vorherige Reviewartikel werden aufgrund der Themenaktualität inkludiert. Abschlussarbeiten sind jedoch von der Suche ausgeschlossen.

Der Anspruch an den Umfang der Literatursuche wurde in Kapitel 3.1 dargelegt. Hierzu bedarf es zunächst einer grundlegenden Recherche zur Identifikation aktueller Arbeiten zu dynamischen MCDA-Methoden. Als Startzeitpunkt der Suche wurde das Jahr 1987 verwendet, da zu diesem Zeitpunkt das Konzept der Dynamik in Gerking (1987) erstmals in Verbindung mit multikriteriellen Entscheidungen in der Literatur thematisiert wurde. Zur Suche wurden folgende Begriffe in den oben genannten Datenbanken verwendet: multi-criteria analysis, multi-criteria decision making, multi-criteria decision analysis, mcda, mcdm, multiple criteria analysis, uncertainty decision, dynamic decision making, dynamic mcda, dynamic multi-criteria, dynamic multiple criteria decision analysis, dynamic mcdm, dynamic mcda. Die gefundenen Arbeiten wurden ferner nach für die Fragestellung geeigneten Referenzen (Rückwärtssuche), als auch nach Zitationen in späteren Arbeiten (Vorwärtssuche) durchsucht. Hierdurch wird trotz fehlender Vollständigkeit eine weitgehende Erfassung bestehender Forschungsstränge ge-

währleistet. Basierend auf diesem Suchprozess wurden nach einer Studie des Titels und Abstracts (sofern vorhanden) 60 Zeitschriftenartikel aus 31 verschiedenen Zeitschriften, 6 Konferenz-Paper, 11 Buchquellen und eine Dissertation gefunden. Die Suche wird als beendet betrachtet, wenn sich trotz fortschreitender Suche keine neuen methodischen Konzepte in den Suchergebnissen erkennen lassen (Webster und Watson (2002)). Mit Abschluss der Suche ist ebenso die zweite methodische Phase beendet.

Gegenstand der nun folgenden dritten methodischen Phase ist eine Verarbeitung der identifizierten Literatur. Diese ist auf Relevanz zu überprüfen, zu verarbeiten und geeignet zu systematisieren (Fettke (2006)).

Die in Phase zwei identifizierten Arbeiten wurden auf Basis einer Lektüre des definitorischen und methodologischen Teils klassifiziert. Als Kategorien zur Klassifizierung wurden, wie in Tabelle 1 dargestellt, das Veröffentlichungsjahr, die Art der Quelle, der Name des Journals, sofern es sich um Artikel aus wissenschaftlichen Zeitungen handelt, das Ranking des Journals (falls vorhanden) und die verwendete Methode gewählt. Ferner wurde untersucht, ob und in welcher Art und Weise eine dynamische Komponente im Sinne der in Kapitel 2.3 gegebenen Definition methodische Berücksichtigung findet. Die vollständige Tabelle kann dem Anhang entnommen werden. Alle Arbeiten ohne Bezug zu dynamischen MCDA-Methoden wurden dabei für eine weitere Berücksichtigung ausgeschlossen.

Anschließend wurde eine Auswahl in dieser Arbeit zu inkludierender Forschungsstränge vorgenommen. Ein Forschungsstrang zeichnet sich, neben einem einheitlichen Verständnis des Dynamikbegriffes, durch aufeinander aufbauende, ähnliche konzeptionelle Ansätze zur Berücksichtigung einer dynamischen Komponente in MCDA-Methoden aus. Zu jedem dieser Forschungsstränge wurden möglichst repräsentative Arbeiten gesucht, anhand welcher eine methodische Entwicklung bis hin zum aktuellen Forschungsstand ersichtlich wird. Beiträge aus höher bewerteten Zeitschriften wurden hierbei denen aus schlechter bewerteten bevorzugt.

4. Einbeziehung der dynamischen Komponente

In diesem Kapitel wird mit der Aufbereitung der Rechercheergebnisse die Analyse aus der vierten methodischen Phase des Reviewprozesses nach Fettke (2006) behandelt, bevor im anschließenden Kapitel fünf die Interpretation erfolgt.

Insgesamt wurden 51 der insgesamt 78 gefundenen Arbeiten als relevant im Sinne der in dieser Arbeit verwendeten Definition von Dynamik erachtet, was einer Quote von 65,4% entspricht. Die gefundenen Arbeiten entstammen den Jahren 1987 bis 2017.

Anhand der in Kapitel 2.3 festgelegten Arbeitsdefinition von Dynamik wurden folgende drei Gruppen für die zu inkludierenden Forschungsstränge gebildet: Szenario-basierte Ansätze, solche, die Lebenszyklus-Modelle verwenden und diejenigen, welche eine dynamische Problemstrukturierung behandeln. Anschließend wurde jede der 51 als relevant er-

achteten Arbeiten genau einer Gruppe zugeordnet, um Redundanzen zu vermeiden. Die ausgefüllte Tabelle der klassifizierten Literatur kann dem Anhang entnommen werden.

Die folgenden drei Unterkapitel betrachten diese drei unterschiedlichen Ansätze, eine dynamische Komponente in multikriterielle Entscheidungsprobleme zu integrieren, nun ausführlich und getrennt voneinander. Die Arbeiten wurden zur Analyse anhand der publizierenden Zeitschrift bzw. Buchtitels, dem entsprechenden Ranking der Zeitschrift sowie der verwendeten MCDA-Methode klassifiziert. Zusätzlich wurden die Arbeiten nach weiteren Methodiken durchsucht, die in Kombination mit MCDA eine Integration der Dynamik ermöglichen. Die Analyse der betrachteten Ansätze erfolgt geordnet nach Forschungssträngen und innerhalb eines Forschungsstrangs jeweils chronologisch. Diese Vorgehensweise wird gezielt gewählt, um die methodischen Entwicklungen bis zu einem möglichst aktuellen Forschungsstand aufzuzeigen und eventuellen weiteren Forschungsbedarf im weiteren Verlauf der Arbeit aufzudecken.

Die als nicht relevant klassifizierten Arbeiten wurden von einer weiteren Betrachtung ausgeschlossen, da sie sich entweder thematisch nicht der MCDA widmen (Crousillat et al. (1993); Millán et al. (1998); Miranda und Proenca (1998); Labuschagne und Brent (2005); Münzberg et al. (2014)), keinen dynamischen Aspekt im Sinne der hier verwendeten Arbeitsdefinition behandeln (Fishburn und Lavalley (1999); Matsatsinis und Samaras (2001); Braune et al. (2009); Kilgour et al. (2010); Mardani et al. (2015b); Rezaei (2015)) oder methodische Lösungen basierend auf Fuzzy MCDA vorschlagen (Mardani et al. (2015a)). Ebenso ausgeschlossen wurden Arbeiten, die sich mit MODM-Methoden beschäftigen (Aouni et al. (2015)). Weitere Arbeiten weisen zwar eine thematische Relevanz auf, wurden jedoch im Rahmen des Auswahlprozesses keinem repräsentativen Forschungsstrang zugeordnet und deshalb nicht weiter analysiert (Stewart (1992); Gorenstin et al. (1993); Linares (2002); Omann (2004); Hites et al. (2006); Mendoza und Martins (2006); Lin et al. (2008); Belton und Stewart (2010); Benítez et al. (2012); Xu und Ouenniche (2012); Jiang et al. (2014); Kluczek (2016); Zolfani et al. (2016a)).

4.1. Szenario-basierte Ansätze

Bei Gerking (1987) findet sich eine der ersten Arbeiten zur Integration von Dynamik in die MCDA von Entscheidungsproblemen mithilfe von Szenarien. Gerking greift das Problem der Unsicherheit bei multikriteriellen Entscheidungsproblemen auf und erweitert dies um eine zeitabhängige, dynamische Komponente. Diese im Zeitverlauf abnehmenden „dynamischen Unsicherheiten“ (engl. dynamic uncertainties, Gerking (1987), S. 193) treten im Zusammenhang mit mehrstufigen Entscheidungsprozessen auf. Diese liegen vor allem bei strategischen Entscheidungen vor, etwa im Bereich des Energiemanagements. Die Besonderheit ist hier unvollkommene Vorhersehbarkeit. Das bedeutet, dass während des mehrstufigen Entscheidungsprozesses unerwartete Änderungen noch zu treffende Entscheidungen beeinflussen können. Gerking's Ansatz berücksichtigt

Tabelle 1: Klassifikation der gefundenen Literatur

Autoren und Veröffentlichungsjahr	Titel	Art der Quelle	Name des Journals	Ranking	Verständnis der dynamischen Komponente	verwendete Methodik
-----------------------------------	-------	----------------	-------------------	---------	--	---------------------

**Abbildung 4:** Analytische Phase des Reviewprozesse

sowohl dynamische Unsicherheiten als auch unvollkommene Vorhersehbarkeit folgendermaßen: Zunächst wird das Entscheidungsproblem abgebildet und eine Lösung für die entscheidungsrelevante Periode erzeugt. Anschließend können in einem Nachbearbeitungsprozess ausgewählte Variablenwerte für zukünftige Perioden verändert werden. Bei einer Fixierung der Variablenwerte für vergangene Perioden können somit schrittweise und für jede Entscheidungsrelevante Variable verschiedene Szenarien abgebildet und ein stufenweiser, dynamischer Entscheidungsprozess ermöglicht werden. Ein aktueller Reviewartikel zum Umgang mit Unsicherheiten wurde von [Stewart und Durbach \(2016\)](#) verfasst, mit einem separaten Kapitel zu MADM-basierten Szenarioplanungsansätzen ([Stewart und Durbach \(2016\)](#), S. 486ff.).

4.1.1. Das MAVT-Rahmenkonzept von Goodwin und Wright

Goodwin und Wright widmen sich in Ihren Arbeiten ebenfalls dem potenziellen Nutzen der Szenarioplanung für strategische Entscheidungen. Dabei gelangen sie zu einer identischen Definition von Szenarien wie Gerking ([Gerking \(1987\)](#)). Ein Szenario ist keine Vorhersage der Zukunft, sondern mehrere Szenarien ermöglichen die Beschreibung plausibler zukünftiger Entwicklungen ([Wright und Goodwin \(1999\)](#)). Zwar hat ein einzelnes Szenario eine bestimmte Eintrittswahrscheinlichkeit, das Ziel ist es jedoch durch eine Spanne an Szenarien den Handlungsrahmen für zu treffende Entscheidungen festzulegen. In [Wright und Goodwin \(1999\)](#) legen die Autoren zunächst unter psychologischen Gesichtspunkten den Nutzen von Szenarioplanung für strategische Entscheidungsträger dar, bevor sie in einer späteren Arbeit die Eignung der Multi-Attribute Value Theory (MAVT), einem klassischen MADM-Verfahren, für die Strukturierung strategischer Entscheidungsprozesse mit Szenariobetrachtung untersuchen ([Goodwin und Wright \(2001\)](#)). Sie kommen dabei zu dem Ergebnis, dass eine MAVT-Modellierung für strategische Entscheidungen mit mehreren Szenarien den Anforderungen entspricht und deutlich besser geeignet ist als eine MAUT-Modellierung (engl. Multi-Attribute Utility Theory), welche sich laut Goodwin und Wright vor allem durch die benötigte Zuweisung von Wahrscheinlichkeiten und die fehlende Transparenz als nicht geeignet erweist. Methodisch sieht das vorgeschlagene MAVT-Rahmenkonzept von Goodwin und Wright folgendermaßen aus:

- Festlegung der strategischen Optionen und Szenarien
- Eine Handlungsalternative entspricht hierbei einer strategischen Option unter einem jeweiligen Szenario
- Festlegung eines Wertebaums, welcher den strategischen Zielen entspricht
- Bewertung der Handlungsalternativen auf Basis des erstellten Wertebaums anhand einer Skala von 0 bis 100
- Gewichtung der Ziele des Wertebaumes mit der ‚swing weight‘-Methode
- Gesamtintegration der Werte für jede Handlungsalternative anhand der im vorherigen Schritt festgelegten Gewichtung

[Montibeller et al. \(2006\)](#) greifen den Konzeptrahmen von [Goodwin und Wright \(2001\)](#) für strategische Entscheidungen auf und diskutieren anhand zweier Fallstudien erstmals die aus diesem Ansatz resultierenden praktische Probleme. Als Erweiterung schlagen Sie die Verwendung von jeweils einem separaten MCDA-Modell für jedes erstellte Szenario vor, um die Gewichtung des Wertebaumes für jedes Szenario unabhängig gestalten zu können und so die Evaluation der strategischen Optionen zu erleichtern. Des Weiteren weisen [Montibeller et al.](#) den Handlungsoptionen eine Robustheit zu und identifizieren über alle Szenarien hinweg eine „sichere“ beziehungsweise robuste Option. Dies geschieht durch die Betrachtung des Wertes der Zufriedenstellung jeder Handlungsoption und die Differenz zum Wert der maximalen Zufriedenstellung von 100. Das geringste Maximum all dieser Differenzen über alle Szenarien repräsentiert dabei die insgesamt robusteste Entscheidung. In einer späteren Arbeit führen [Ram und Montibeller \(2013\)](#) diesen Entscheidungsfindungsprozess unter dem Gesichtspunkt der Robustheit fort und wenden es zudem in der Praxis an. Zukünftigen Forschungsbedarf sehen die Autoren in einem ausführlichen Vergleich zwischen MCDA in Kombination mit Szenarioplanung und herkömmlichen MCDA-Methoden basierend auf Entscheidungsbäumen sowie in einer Optimierung der Methode für komplexe strategische Entscheidungen mit vielen Handlungsoptionen.

Durbach und Stewart betrachten in Ihrem Review unter anderem auch die Konstruktion von Szenarios für den Um-

gang mit Unsicherheit bei komplexen strategischen Entscheidungen (Durbach und Stewart (2012), S. 8 ff.) und kommen zu dem Ergebnis, dass Szenario-basierte MCDA-Ansätze eine vielversprechende Möglichkeit für den Umgang mit Unsicherheit darstellen, wenngleich dieses Forschungsfeld jedoch noch weiterer Anwendungsarbeiten bedarf. Zur Arbeit von Montibeller et al. (2006) führen sie an, dass eine Argumentation mithilfe der Robustheit jedoch auch irreführend sein kann.

In Stewart et al. (2013) wird diese Kritik am Ansatz von Montibeller et al. (2006) anhand eines numerischen Beispiels konkretisiert. Ferner kommen Stewart et al. zu dem Schluss, dass MAVT-Methoden aber auch andere Methoden der quantitativen Entscheidungsanalyse, welche keine Nutzenfunktionen und Wahrscheinlichkeitsfunktionen von korrelierenden Variablen verlangen, gleichermaßen geeignet erscheinen. Als Richtlinie für strategische Entscheidungen schlagen sie die Generierung von 4-6 Szenarien vor (Stewart et al. (2013), S. 683). Des Weiteren werden Metakriterien, eine Kombination aus Szenarien und primären Kriterien, eingeführt. Diese ermöglichen es, jedes Kriterium in mehreren Szenarien gleichzeitig zu betrachten.

In Goodwin und Wright (2014) findet sich eine Anwendung von Entscheidungsunterstützung für strategische Verhandlungen der Managementebene. Auf Basis der MAVT können hierdurch die Werte und Ziele von Verhandlungspartnern antizipiert werden, ohne diese direkt von ihnen zu erheben und auf diese Weise eine vorteilhafte Verhandlungsposition zu schaffen. Eine methodische Weiterentwicklung dieses Ansatzes schlagen Cairns et al. (2016) in einer anschließenden Arbeit mit Fokus auf den Wechselwirkungen von mehreren Stakeholdern mit unterschiedlichen Zielen und Werten vor. Schrittweise wird durch die Verwendung der CSR-Methode (engl. critical scenario method) jeder Alternativen-Stakeholder-Kombination für jedes Szenario ein Zufriedenstellungswert zugeordnet und anschließend auf Basis der Verhandlungsmacht der jeweiligen Gruppe von Stakeholdern gewichtet.

4.1.2. Anwendungen und Erweiterungen des Rahmenkonzeptes von Goodwin und Wright

Basierend auf den vorherigen Arbeiten (Goodwin und Wright (2001); Montibeller et al. (2006); Stewart et al. (2013)) schlagen Scholten et al. (2014) einen neuen Ansatz für robuste und nachhaltige Sanierungsstrategien im Wassermanagement vor. Hierzu kombinieren sie Methoden des Vermögensmanagements, der Schadensmodellierung, der Entscheidungsunterstützung sowie der Szenarioanalyse. Die Verwendung von MCDA bzw. einem MAVT/MAUT-Modell ermöglicht laut Scholten et al. das Finden einer robusten und nachvollziehbaren strategischen Entscheidung.

In Lienert et al. (2015) präsentieren die Autoren ausführlich einen praktischen Problemstrukturierungsprozess mithilfe von Szenarioplanung und MAUT-Analyse mit besonderem Fokus auf der Bestimmung des Problemkontexts, der Stakeholderauswahl, Szenariodefinition und Zielhierarchie samt

Attributbestimmung. Die anschließenden Schritte der Bestimmung der Konsequenzen für jedes Szenario, der Bestimmung der Präferenzen und Gewichtungen, Durchführung einer Sensitivitätsanalyse zum Identifizieren der robusten Alternativen sowie das Einholen von Stakeholder-Feedback werden in Zheng et al. (2016) behandelt. Zusätzliche Informationen zur durchgeführten MAUT-Analyse, insbesondere zu einer globalen Sensitivitätsanalyse anstelle einer lokalen findet sich in Scholten et al. (2015). Aufgrund der einfacheren Bestimmung von Wertefunktionen im vorliegenden Problemkontext wird zunächst diese gebildet und anschließend in eine Nutzenfunktion transformiert. Eine ausführliche Beschreibung des Umwandlungsprozesses liefert die Arbeit von Reichert et al. (2015).

Einer erweiterten Aggregationsfunktion widmet sich Zahir (2016) in seiner Arbeit, um unerwünschte Vertauschungen, verursacht durch das Entfernen oder Hinzufügen von Alternativen nach bereits abgeschlossener Normalisierung, zu vermeiden. Hierzu schlägt er eine modifizierte gewichtete Summenregel für die Anwendung in AHP (engl. analytic hierarchical process) und MAVT-Methoden vor und liefert ein numerisches Anwendungsbeispiel.

4.1.3. Weitere Ansätze zur Integration einer dynamischen Komponente in MCDA-Methoden

Die im vorherigen besprochenen Arbeiten greifen sehr stark auf die klassischen MADM-Verfahren der amerikanischen Schule zurück. Hingegen verwenden Vaillancourt und Waaub (2004) anstelle der MAVT bzw. MAUT mit PROMETHEE einen Outranking-Ansatz der europäischen Schule zur Untersuchung einer gerechten Beschränkung von Treibhausgasemissionen auf internationalem Level. Das vorgeschlagene MCDA-Modell berücksichtigt eine dynamische Komponente im Sinne einer veränderbaren Kriteriengewichtung und in der anschließenden Evaluation durch die Einführung veränderbarer Prioritäten, um die weltweit nicht einheitliche Auffassung bezüglich einer gerechten Zuweisung von Emissionsbeschränkungen abbilden zu können. Weitere Methoden und Anwendungen von PROMETHEE werden ausführlich im Reviewbeitrag von Behzadian et al. (2010) aufgeführt und besprochen.

Eine aktuellere Arbeit von Schmuck et al. (2012) zeigt die Verwendung von PROMETHEE zur Wahl eines Energieerzeugungssystems in der Gemeinde Ludwigsfelde. Hierzu wurden sieben Energieversorgungs-szenarien erstellt und die Kriteriengewichtung wurde mithilfe einer „stillen Abstimmung“⁵. Wilkens und Schmuck sehen den Vorteil von Outranking-Ansätzen in der geringeren Kompensation während der Aggregation im Vergleich zu nutzen- und wertebasierten Methoden, was für den Einsatz im Nachhaltigkeitsmanagement wünschenswert sei.

⁵engl. silent negotiation. Hierbei teilt jeder Entscheidungsträger seine Gewichtung in Form einer Abstimmungskarte mit. Auf diese Weise soll eine, bedingt durch die große Anzahl an Entscheidungsträgern, zu ausführliche Diskussion vermieden und eine demokratische Stimmgleichheit gewährleistet werden (Schmuck et al. (2012), S. 615 nach; Pictet und Bollinger (2005))

Ein weiteres Anwendungsbeispiel für die Wahl eines Energieversorgungssystems am Beispiel Portugals von Ribeiro et al. (2013) generiert zunächst fünf verschiedene Szenarien für zukünftige Energieversorgungssysteme mithilfe eines gemischt-ganzzahligen linearen Optimierungsmodells⁶ in GAMS und evaluiert diese anschließend separat in einem MCDA-Tool mit additiver Wertefunktion.

Frini (2014) knüpft thematisch an die Arbeit von Ribeiro et al. an und präsentiert ein multikriterielles Entscheidungsunterstützungsmodell mit mehreren Szenarien und einer Evaluation über mehrere Perioden zur Auswahl eines nachhaltigen Energieversorgungssystems. Durch die Evaluation von mehr als nur einer zukünftigen Periode soll das Finden einer nachhaltigen Lösung verbessert werden. Hierzu schlägt Frini ein vierstufiges Modell vor. Innerhalb der Phasen kommen dabei unter anderem verschiedene MADM-Methoden zum Einsatz: In der Phase der Problemstrukturierung und Unsicherheitsbewertung die Methoden AHP und MACBETH. In den anschließenden Aggregationsphasen wird über die vorher festgelegten Kriterien, Perioden und Szenarien jeweils die Anwendung von TOPSIS aufgrund der geringen Anforderungen an die Präferenzmodellierung vorgeschlagen, um zur finalen Auswertung zu gelangen.

Auffallend bei Arbeiten der letzten 15 Jahre zum Anwendungsgebiet der nachhaltigen Entwicklung ist neben der methodischen Integration von Szenarioplanung und multikriterieller Entscheidungsunterstützung der häufige Einsatz von Lebenszyklusmodellen. So unter anderem in der hier besprochenen Arbeit von Wilkens und Schmuck. Deshalb wird das folgende Unterkapitel diese methodische Komponente genauer betrachten.

4.2. Verwendung von Lebenszyklus-Modellen

Neben der im vorherigen Abschnitt adressierten Unsicherheit bezüglich der Entwicklungen externer Einflussfaktoren weisen Energieversorgungssysteme sowie Technologien im Allgemeinen ebenso eine zeitabhängige Dynamik auf und unterliegen keinesfalls einer stetigen und absehbaren Entwicklung. Während neue Technologien in ihrer Umweltfreundlichkeit den etablierten zwar zunächst unterlegen scheinen und ihre Einführung zunächst mit Risiken und Unsicherheiten verbunden ist, können sie doch das Potential für eine zukünftig größere Leistungsfähigkeit mit sich bringen. Als Instrument zur Beurteilung von Produkten und Prozessen wurde deshalb das sogenannte LCA (engl. Life-cycle assessment) entwickelt (Geldermann und Rentz (2005); nach Guinée (2002)). Für Entscheidungen im Kontext nachhaltiger Entwicklungen erscheint aufgrund der hohen Konflikträchtigkeit des Themas zusätzlich ein Einsatz der MCDA als geeignet, wie Munda in seiner Arbeit ausführt (Munda (2016)). Dieses Kapitel soll Ansätze untersuchen, welche diese beiden Instrumente miteinander kombinieren.

⁶engl. Mixed Integer Linear Programming, kurz MILP; für weitere Erläuterungen siehe (Kallrath (2013))

Ein erstes Rahmenkonzept, bestehend aus integriertem Lebenszyklus- und Entscheidungsmodell mit Berücksichtigung einer dynamischen Komponente in Form eines langfristigen Zeithorizontes bei der Energiesystemplanung, liefern Polatidis et al. (2003). In einer späteren Arbeit werden Outranking, Wert- beziehungsweise Nutzenfunktions-basierte sowie einige weitere MCDA-Methoden auf einen Einsatz innerhalb des vorgestellten Rahmenkonzeptes vergleichend überprüft (Polatidis et al. (2006)).

Eine weitere vergleichende Betrachtung der Anwendung von MAUT, AHP, ELECTRE und PROMETHEE in der Nachhaltigkeitsbewertung auf Basis zehn verschiedener Vergleichskriterien führen Cinelli et al. (2013) in einer späteren Arbeit durch. Die Ergebnisse können der Tabelle 4.2 entnommen werden:

4.2.1. Klassische Ansätze zur Integration von MCDA in die Lebenszyklus-Analyse

Eine mögliche Anwendung von AHP zeigen Hermann et al. (2007) mit der Entwicklung der Methode COMPLIMENT auf. COMPLIMENT (engl. COMbining environmental Performance indicators, Life cycle approach and Multi-criteria to assess the overall ENvironmental impacT) kombiniert die Stärken und Einsatzwecke der Instrumente der LCA, MCDA sowie von Leistungskennzahlen eines Unternehmens bezüglich Umweltfreundlichkeit⁷ in einer gemeinsamen Methode, um eine genauere Bewertung von Umweltbelastungen bestimmter Geschäftsbereiche oder Prozesse anhand eines Indexes zu erhalten. Der Nutzen von AHP liegt hierbei in der Anwendung im abschließenden Aggregationsprozess auf die durch EPI erhaltenen und von LCA klassifizierten Daten bezüglich der Umweltbelastung. Ebenso wie Hermann et al. schlagen Halog und Manik (2011) eine Verwendung von AHP in einem integrierten Rahmenkonzept für die Nachhaltigkeitsbewertung betrieblicher Aktivitäten vor. Die Aufgabe der MCDA liegt hierbei in der Berücksichtigung von Ansprüchen der verschiedenen Stakeholder. Anschließend wird durch Data Envelopment Analysis (DEA), Agent Based Modelling (ABM), System Dynamics (SD), Sustainable Network Theory (SNT) sowie Szenarioplanung die durch Interdependenzen und Unsicherheiten entstehende Dynamik methodisch verarbeitet. Die Autoren präsentieren hiermit eine der wenigen Arbeiten, welche eine Applikation von LCA und Input-Output-Analyse methodisch verbindet. Eine Forschungslücke, die auch Onat et al. (2016) in einem kurzen Reviewkapitel Ihrer Arbeit aufzeigen. Desweiteren sehen Onat et al. weiteren Forschungsbedarf bei der Integration von MODM und LCA, was jedoch aufgrund der in dieser Arbeit getroffenen Fokussierung auf MADM-Methoden nicht weiter ausgeführt werden soll. Einen jüngeren Ansatz zur Integration von MADM-Methoden der amerikanischen Schule in LCA steuern Scott et al. (2016) bei. Sie wenden die MAUT auf die Ergebnisse der LCA über mehrere Szenarien hinweg an, um

⁷für nähere Informationen zu Environmental Performance Indicators, kurz EPI, siehe (Jasch (2000))

Tabelle 2: Ergebnisse der Untersuchung von Cinelli et al.; Quelle: [eigene Darstellung nach Cinelli et al. (2013)]

Evaluationskriterium	MAUT	AHP	ELECTRE	PROMETHEE
Lebenszyklus-Perspektive	möglich	möglich	möglich	möglich
Anfälligkeit für Kompensation	hoch	hoch	nicht möglich/ teilweise	teilweise/ hoch
Wahl der Gewichtungsfaktoren	eingeschränkt	eingeschränkt/ alle wichtigen Koeffizienten	alle wichtigen Koeffizienten	eingeschränkt/ alle wichtigen Koeffizienten
Einbeziehung verschiedener Schwellenwerte	nicht möglich/ möglich	nicht möglich	möglich	möglich
qualitative und quantitative Daten	möglich	möglich	möglich	möglich
Sensitivitätsanalyse/ Unsicherheitsbewertung	möglich	teilweise möglich/ möglich	möglich	teilweise möglich/ möglich
Robustheit	Rangumkehrung nicht möglich	Rangumkehrung möglich	Rangumkehrung möglich	Rangumkehrung möglich
Benutzerfreundlichkeit	hoch/ gering	hoch/ mittel	gering	mittel
Lernfähigkeit des Modells	schwierig implementierbar	schwierig implementierbar	schwierig implementierbar	einfach implementierbar mittels Szenarioanalyse
Graphische Veranschaulichung	großes Angebot vorhanden	großes Angebot vorhanden	großes Angebot wenig Möglichkeiten zur graphischen Veranschaulichung	großes Angebot an Software vorhanden

Unsicherheiten, bedingt durch Umwelteinflüsse, Modellannahmen sowie durch von Entscheidungsträgern festgelegte Präferenzen, besser abschätzen zu können. Laut Scott et al. liegt der Vorteil von MAUT hierbei vor allem in der hohen Transparenz und der Möglichkeit, die Präferenzen der Entscheidungsträger über eine Nutzenfunktion einfließen zu lassen.

4.2.2. Anwendungsbeispiele und Entwicklung von Outranking-Methoden

Geldermann und Rentz (2005) präsentieren eine Möglichkeit, das Instrument der LCA mit der MCDA zu verknüpfen und veranschaulichen dies anhand einer Fallstudie aus dem Sektor der industriellen Lackierung. Hierbei wenden sie den Outranking-Ansatz PROMETHEE aus der europäischen Schule an und vergleichen die Ergebnisse mit einer Anwendung von MAUT und AHP. Sie kommen zunächst zu weitestgehend konsistenten Ergebnissen, wenngleich auch weiterer Forschungsbedarf bezüglich der Vor- und Nachteile der verschiedenen MCDA-Methoden aufgezeigt wird.

5. Einbeziehung der dynamischen Komponente

Der Ansatz von Oberschmidt et al. (2010) zielt darauf ab, die unterschiedlichen Energietechnologien der jeweiligen Entwicklungsphase zuzuordnen und somit individuell gewichten zu können. Ferner wird die Anwendung des Ansatzes anhand einer Fallstudie für das Bioenergiedorf Jühnde veranschaulicht. Die Fallstudie vergleicht vier alternative Möglichkeiten zur Elektrizitätserzeugung miteinander und teilt die Technologien und Kriterien in die vier Lebenszyklen „Einführungsphase“, „Wachstumsphase“, „Reifephase“ und „Sättigungsphase“ ein. Als MCDA-Methode wird sowohl PROMETHEE I als auch PROMETHEE II verwendet. Das Ergebnis wird in Form eines relativen Vergleichs der vier zur Verfügung stehenden Alternativen präsentiert. Ein ausführliches Review zum Einsatz von MCDA-Methoden bei der Bewertung von Bionergiesystemen findet sich ergänzend dazu bei Scott et al. (2012). Hingegen versuchen Domingues et al. (2015) in einer späteren Arbeit, die leichte Nutzfahrzeuge anhand ihrer Umweltbelastung bewertet, unter Verwendung von ELECTRE TRI, anders als Oberschmidt et al. (2010), eine absolute Evaluation jeder Alternative zu erhalten.

5.0.1. Weitere Integrationsmöglichkeiten von Lebenszyklusmodellen und MCDA

Im Rahmen der Bewertung von Umweltbelastungen bei der Biomassenerzeugung, setzen Myllyviita et al. (2012) gezielt das SMART-Verfahren⁸ in der Gewichtungphase der LCA ein, um die verschiedenen, im Produktionsprozess auftretenden, Umweltauswirkungen unterschiedlich stark in

die Evaluation einfließen zu lassen. Insgesamt bewerten die Autoren einen integrierten Ansatz von MCDA und LCA als geeignet, um dynamischen Unsicherheiten im Sinne der hier verwendeten Arbeitsdefinition zu begegnen, jedoch sehen sie noch weiteren Forschungsbedarf bezüglich der weiteren Integration von Werkzeugen der MCDA, beispielsweise zur Normalisierung von Daten.

Neben der Beurteilung von Auswirkungen auf die Umwelt mithilfe der LCA lassen sich sozioökonomische Auswirkungen wirtschaftlicher Aktivitäten gleichermaßen untersuchen. Dieses erweiterte Konzept findet sich etwa bei Macombe et al. (2013) unter dem Begriff SLCA (engl. Social Life Cycle Assessment). Neben dem Erfassen des State-of-the-Art der Lebenszyklusanalyse, evaluieren Macombe et al. im Rahmen ihrer Arbeit die sozioökonomischen Auswirkungen von Biodieselproduktion und erfassen somit eine weitere dynamische Unsicherheit methodisch. Wie auch Myllyviita et al. (2012) erachten sie die Verwendung von MCDA als methodischen Rahmen für die LCA beziehungsweise SLCA generell als geeignet, sehen jedoch noch Ausbesserungsbedarf in der methodischen Gestaltung dieser Integration.

5.1. Dynamische Problemstrukturierung

Die bislang vorgestellten Ansätze erfassen die dynamische Komponente über eine Kombination von weiteren methodischen Werkzeugen mit den Methoden der MCDA. Für eine Erweiterung der herkömmlichen MCDA, etwa durch die Verwendung mehrerer Datensätze zu Alternativen und Kriterien über einen gewissen Zeitraum und deren methodische Einbeziehung in den Entscheidungsprozess, hat sich in der Literatur der Begriff der DMCDa entwickelt (Chen et al. (2010)). Die direkte Integration von dynamischen Aspekten in diesen DMCDa-Methoden und die methodische Entwicklung dieses Forschungsfeldes soll nun in diesem Abschnitt an repräsentativen Arbeiten aufgezeigt werden.

5.1.1. Konzeptuelle Entwicklungen zu DMCDa-Modellen

Kornbluth (1992) präsentiert in seiner Arbeit einen empirischen Befund dafür, dass das Verändern der Kriteriengewichtung in einem mehrstufigen Entscheidungsprozess, bedingt durch äußere Entwicklungen, langfristig bessere Ergebnisse bezüglich der Entscheidungsqualität liefern kann.

Aufbauend auf der Arbeit von Henig und Buchanan (1996) entwickeln Corner et al. (2001) ein Konzept für eine dynamische Gestaltung des Problemstrukturierungsprozesses. Hiernach sollen sich Kriterien und Alternativen im Laufe des Prozesses jeweils gegenseitig generieren, anstelle einer statischen und getrennt voneinander stattfindenden Definition dieser beiden Elemente. Als Startpunkt kann im Rahmen dieser dynamischen Problemstrukturierung sowohl mit der Bestimmung von Alternativen über eine Alternativenfokussierte Denkweise (engl. Alternative-Focused Thinking, AFT), als auch mit der Festlegung von Kriterien durch eine Werte-orientierte Denkweise (engl. Value-Focused Thinking) begonnen werden.

Zusätzlich liefern Corner et al. noch ein illustratives Beispiel ihres Ansatzes und empirische Belege anhand von

⁸Methode zur Kriteriengewichtung auf Basis von vorher vergebenen Punkten als Orientierungswert und gewichteten linearen Durchschnitten. Für weitere Erläuterungen siehe (Von Winterfeldt und Edwards 1986 Kapitel 8.2)

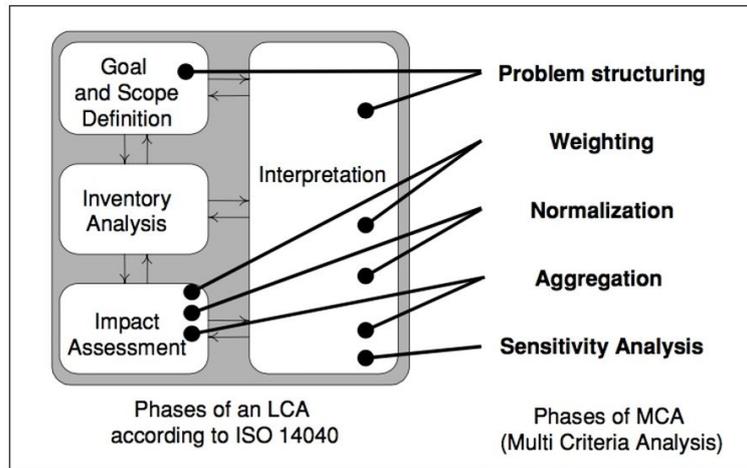


Abbildung 5: Verfahrensschritte der LCA und MCDA; Quelle: [Geldermann und Rentz (2005)]

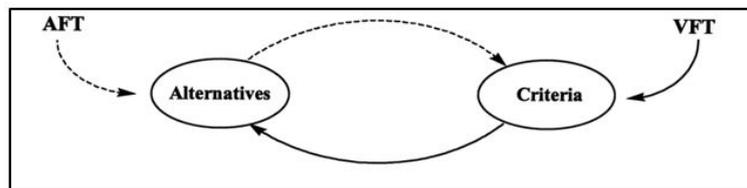


Abbildung 6: Dynamischer Problemstrukturierungsprozess nach Corner et al., Quelle: [Corner et al. (2001)]

Fallstudien und entscheidungstheoretischer Literatur. Saaty (2007) stellt drei verschiedene Herangehensweisen an zeitabhängige und somit dynamische Entscheidungen mithilfe des analytischen Hierarchieprozesses (AHP) vor und erläutert die mathematischen Hintergründe und Zusammenhänge der funktionalen Herangehensweise ausführlich:

1. Über die Integration von Szenarios mit klar definierten Zeitabschnitten in das Entscheidungsmodell und anschließend paarweisen Vergleichen für jeden Zeitabschnitt (strukturelle Herangehensweise)
2. Über paarweise Vergleiche zu verschiedenen Zeitpunkten und dem Aufzeigen der Entwicklung der relativen Änderungsrate im Zeitverlauf für das Vergleichspaar. Hieraus können zeitabhängige Urteile und Prioritäten abgeleitet werden (funktionale Herangehensweise)
3. Über paarweise Vergleiche anhand von mathematischen Funktionen und der Ableitung einer dynamischen Funktion aus diesen (hybride Herangehensweise)

5.1.2. DMCD-Modelle basierend auf Feedback-Daten

In Pais und Ribeiro (2009) präsentieren die Autoren eine mögliche Architektur für die Gestaltung von DMCD-Problem auf Feedback-Basis. Die Besonderheit dieses iterativen Vorgehens liegt dabei im vorgeschlagenen dynamischen Evaluationsprozess. Hierbei wird ein historischer Datensatz zu Alternativenbewertungen aus der vorherigen Iteration mit dem Ergebnis der aktuellen Bewertungsprozedur aggregiert und somit nach jeder Iteration eine aktuelle Rangfolge an

Alternativen ausgegeben. Die Bewertungsprozedur basieren Pais und Ribeiro auf der Methode der einfachen additiven Gewichtung (engl. Weighted Sum Model, WSM).

Yu und Chen (2010) gelangen durch Integration von MCDA mit den Feldern der neuronalen Informationsverarbeitung und Psychologie zu einer Notwendigkeit von entscheidungstheoretischen Modellen mit veränderbaren Strukturelementen, liefern jedoch kein mathematisches Rahmenkonzept für ein solches DMCD-Modell. Dieses findet sich unter anderem bei Chen et al. (2010) anhand eines Beispiels aus dem medizinischen Notfallmanagement. Die Erweiterung zum ebenfalls präsentierten klassischen MCDA-Rahmenkonzept liegt in der zusätzlichen Verwendung von Entscheidungstabellen mit Daten vergangener Perioden im Entscheidungsprozess. Für die Gestaltung des Aggregationsprozesses werden drei Möglichkeiten aufgezeigt: mit Präferenz für neuere [ältere] Datensätze oder der Angabe eines Referenzzeitpunktes, um den die Präferenzen normalverteilt sind. Anschließend an diese Aggregation der Feedback-Daten schlagen Yu und Chen nach der Festlegung der Kriteriengewichtung eine auf DEA (engl. Data envelopment analysis) basierende Gesamtintegration zur Erlangung eines Ergebnisvektors für alle Alternativen vor. Campanella und Ribeiro (2011) präsentieren anknüpfend an die Arbeit von Yu und Chen einen Konzeptrahmen für DMCD mit einem numerischen Beispiel der Wahl eines Helikopterlandeplatzes. Beim vorgeschlagenen Rahmenkonzept handelt es sich, ähnlich zu Pais und Ribeiro (2009), um einen iterativen Prozess mit Verwendung einer „Memory“-Funktion in Form historischer

Datensätze. Ergänzend hierzu werden Richtlinien zur Datenspeicherung für die historischen Datensätze diskutiert, um sowohl eine gute Praktikabilität als auch eine möglichst hohe Ergebnisqualität zu erlangen. Das Abbruchkriterium sollte laut Campanella und Ribeiro je nach Problemkontext und -spezifität individuell gewählt werden. Mögliche Einsatzgebiete dieser DMCDAs sehen die Autoren im Notfallmanagement, der Patientendiagnose, der Luft- und Raumfahrt sowie bei der Lieferantwahl. Zu letzterem Anwendungsgebiet steuern Campanella et al. (2011) unter Verwendung des hier vorgestellten DMCDAs-Rahmenkonzeptes aus Campanella und Ribeiro (2011) eine ausführliche Fallstudie bei.

Zulueta et al. (2013) erweitern den Ansatz von Campanella und Ribeiro um einen selektiven und dynamischen Index, um für den Fall einer gleichen Bewertung zweier Alternativen die zeitliche Entwicklung der dynamischen Bewertung in die Entscheidung einzubeziehen und somit auch bei Indifferenz eine fundierte Entscheidung zu ermöglichen. Hierzu schlagen Zulueta et al. die Verwendung bipolarer Skalen im Aggregationsvorgang zur Beschreibung der Entwicklung der Alternativenbewertung im Zeitverlauf vor. Ein Wert von 0 impliziert hierbei keine Veränderung zwischen zwei aufeinanderfolgenden Perioden und ein Wert von 1 [-1] jeweils die größte normalisierte Verbesserung [Verschlechterung].

5.1.3. DMCDAs-Modelle basierend auf Feedback- und Feedforward-Daten

Eine methodische Weiterentwicklung des Ansatzes von Campanella und Ribeiro (2011) um taktische Informationen bzw. Feedforward-Daten schlagen Jassbi et al. (2014) vor. Diese können beispielsweise auf Expertenschätzungen oder Vorhersagemodellen beruhen und bei taktischen und strategischen Geschäftsentscheidungen wie der Lieferantwahl neben historischen Daten in den Prozess der Entscheidungsfindung einbezogen werden. Für eine ausführliche Arbeit zu MCDA-Methoden für die Lieferantwahl sei an dieser Stelle auf den Reviewartikel von Ho et al. (2010) verwiesen. Die methodische Umsetzung dieses zusätzlichen Datensatzes erfolgt mithilfe einer weiteren Matrix zusätzlich zu den beiden bereits bestehenden Matrizen für die historischen Daten und der aktuellen Entscheidungstabelle, wie Abbildung 7 verdeutlicht. Zu jeder Iteration können diese Matrizen zusätzlich verändert werden. Hierdurch lassen sich historische Daten aktualisieren, verbessern und eine stets aktuelle Prognose berücksichtigen. Des Weiteren können bei jeder Iteration Alternativen und Kriterien verändert werden, was einen dynamischen und flexiblen Entscheidungsprozess garantiert.

Eine noch stärkeren Fokus auf zukünftige Datensätze legen Zolfani et al. (2016a) mit der Einführung eines Rahmenkonzeptes für PMADM-Modelle (engl. Prospective Multi-Attribute Decision Making). Hierbei wird die Entscheidungsmatrix neben Kriteriengewichtungen zusätzlich um das Setzen von Limitierungswerten und deren Eintrittswahrscheinlichkeit ergänzt. Diese Limitierungswerte repräsentieren äußere Einflüsse auf die Alternativenbewertungen, welche unabhängig von den Werten des Entscheidungsträgers

sind, aber dennoch zukünftig die Entscheidung beeinflussen könnten. Der Hauptunterschied zu Szenarien liegt laut Zolfani et al. darin, dass die bestmögliche Alternative keine Limitierung aufweist und Limitierungen somit eher als negative Szenarien zu verstehen sind, während Szenarien eine Spanne von möglichen Entwicklungen abdecken.

5.1.4. Dynamische MCDA-Ansätze mit TOPSIS

Im Rahmen der Suche nach einem geeigneten Modell zur Auswahl von nachhaltigen und umweltfreundlichen Lieferantenbeziehungen diskutieren Orji und Wei (2015) den kombinierten Einsatz von System Dynamics und TOPSIS mit Alternativenbewertungen für vergangene, gegenwärtige und zukünftige Perioden. Hierbei wird TOPSIS zur Berechnung der normalisierten Punktwerte der Alternativen bzw. Lieferanten für einen entsprechenden Zeitraum verwendet und die hieraus erlangten Werte anschließend als Input für das auf System Dynamics basierende Selektionsmodell genutzt.

Anknüpfend an die DMCDAs-Methoden von Chen et al. (2010), Campanella und Ribeiro (2011) sowie Jassbi et al. (2014) entwickeln Watróbski et al. (2016) ein dynamisches Entscheidungsunterstützungsmodell unter der gleichzeitigen Verwendung einer Fuzzy TOPSIS sowie einer deterministischen TOPSIS-Methode zur Datenaggregation. Als Aggregationsstrategien wurden zum einen eine gleiche Gewichtung aller Zeitperioden und zum anderen eine Präferenz für aktuellere Perioden, bekannt aus Chen et al. (2010), gewählt. Watróbski et al. kommen zu dem Fazit, dass das untersuchte DMCDAs-Modell durch die Verwendung von Fuzzy TOPSIS und mehreren Aggregationsstrategien eine hohe Praktikabilität und Flexibilität bietet. Die Autoren sehen vor allem in der Untersuchung weiterer Aggregationsstrategien und einer damit verbundenen Modellabänderung weiteren Forschungsbedarf.

6. Vergleich der ausgewählten multikriteriellen Entscheidungsunterstützungssysteme

Ausgehend von den im vorhergehenden Abschnitt berichteten Befunden der Literaturanalyse werden nachfolgend die erzielten Ergebnisse näher diskutiert und damit gleichzeitig die vierte methodische Phase nach dem Schema von Fettke abgeschlossen.

Hierzu wird die in Kapitel vier vorgestellte Literatur auf Auffälligkeiten und Muster hin untersucht und im Hinblick auf den Beitrag zur Forschungsfrage kritisch bewertet. Auch die bei der Recherche verwendete Suchmethodik soll nachträglich durch das Aufzeigen von besonderen Auffälligkeiten und Hindernissen während des Suchprozesses kurz besprochen und bewertet werden, um Hinweise für zukünftige Arbeiten zu liefern.

Ziel dieses Kapitels ist es, durch Diskussion bisherige Erkenntnisse und eventuelle Forschungslücken auf dem Feld der dynamischen multikriteriellen Entscheidungsunterstützung aufzuzeigen, um im resümierenden Kapitel sechs zu einer Beantwortung der anfangs formulierten Forschungsfragen zu gelangen.

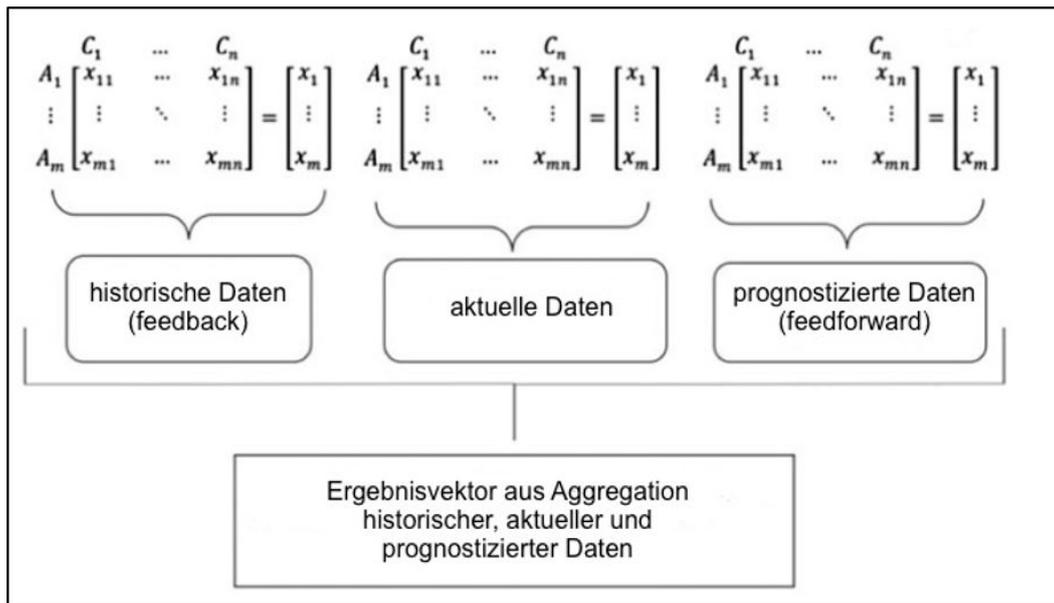


Abbildung 7: Erweitertes DMCDA-Modell von Jassbi et al.; Quelle: [eigene Darstellung nach Jassbi et al. (2014)]

6.1. Kritik bestehender Ansätze

Wie die Analyse der nach dem Suchprozess als relevant klassifizierten Literatur in Kapitel vier gezeigt hat, existieren verschiedene Möglichkeiten zur Integration von Dynamik in Form einer zeitlichen Komponente in MCDA-Methoden. Im Rahmen dieses Reviewartikels wurde hierbei auf drei Herangehensweisen besonderer Fokus gelegt, um eine methodische Entwicklung anhand repräsentativer Arbeiten aufzuzeigen.

6.1.1. Szenario-basierte Ansätze

Die Entwicklung von MCDA-Ansätzen unter methodischem Einbezug von Szenarien beginnt im Jahr 1999 mit der Arbeit von Goodwin und Wright. Frühere Arbeiten zu konkreten Methoden mit Anwendungsbeispielen und Fallstudienberichten wurden im Rahmen der Literatursuche nicht identifiziert. Das Hauptziel von Szenario-basierten Ansätzen ist die Bestimmung der Auswirkung einer Alternativenauswahl innerhalb eines bestimmten Szenarios und der Vergleich mit den weiteren Szenario-Handlungsalternativen-Kombinationen. Aufgrund der Annahme, dass sich die Präferenzen der Entscheidungsträger über alle Szenarien hinweg stetig und konstant verhalten, fokussiert der Ansatz von Goodwin und Wright sich hierbei auf die Ergebnisse innerhalb der Szenarien und strebt keine globale Evaluation an. Hingegen gehen die späteren Weiterentwicklungen von Montibeller et al. (2006) und Stewart et al. (2013) explizit von unterschiedlichen Präferenzen für jedes Szenario aus. Diese Annahme erscheint vor allem im Anwendungsgebiet strategischer Entscheidungen sinnvoller, da externe Unsicherheiten und zeitliche Dynamiken somit flexibler berücksichtigt werden können. Legt ein Entscheidungsträger im Falle sehr guter Konjunkturentwicklungen geringere Präferenz auf kostenbezogene Kriterien, kann sich dies im Rezessions-Szenario entscheidend ändern (Scholten et al. (2014)).

Stewart und Durbach (2016) bewerten eine Integration von MCDA und Szenarioplanung insgesamt jedoch als leistungsfähiges Instrument mit dem Vorteil einer hohen Transparenz für die Entscheidungsträger. Ebenfalls sehen sie es als vorteilhaft an, dass sämtliche MCDA-Methoden hierbei ihre Verwendung finden können. Klärungsbedarf bestehe jedoch noch hinsichtlich der Anzahl festzulegender Szenarien und deren Konstruktionsweise. Als methodisch sinnvoller und transparenter Ansatz für eine möglichst objektive Szenariogenerierung erscheint der von Ribeiro et al. (2013) unterbreitete Vorschlag, welcher durch Formulierung von MILP-Modellen und anschließender Lösung in einer MILP-solver Software Szenarien generiert.

Nach Zolfani et al. (2016b) verwenden die bisher präsentierten Ansätze MADM-Methoden hauptsächlich als unterstützende Methode für die Szenarioplanung- und Auswertung. Die Autoren sprechen in diesem Zusammenhang von Szenario-basierten MADM-Methoden und schlagen gleichzeitig für die Entscheidungsfindung in einem dynamischen Umfeld ein MADM-basiertes Rahmenkonzept vor. Dieses versteht Szenarien lediglich als einen Teil des Entscheidungsprozesses zur besseren Erfassung der Zukunft, lässt jedoch generell auch andere Methoden als die Szenarioplanung für diesen Zweck zu.

6.1.2. Kombination von MCDA und Lebenszyklusanalyse

Während die in dieser Arbeit vorgestellte Literatur zu Szenario-basierten Ansätzen für strategische Entscheidungen auf verschiedensten Themengebieten vorgeschlagen wurde, beschränkt sich das Anwendungsfeld der in Kapitel 4.2 präsentierten Arbeiten vor allem auf die Gebiete der nachhaltigen Entwicklung und Technologiebewertung. Zudem kann die LCA als zusätzliches methodisches Werkzeug lediglich die Dynamik von Technologien im Zeitverlauf erfassen. Für

eine vollständige Erfassung zusätzlicher dynamischer Komponenten im Sinne zeitlich bedingter äußerer Einflüsse ist zusätzlich die Verwendung weiterer Werkzeuge, wie etwa der Szenarioanalyse, nötig (Cinelli et al. (2013)).

Weiterhin ist bei der Beurteilung nachhaltiger Entscheidungen anzumerken, dass lediglich gut erforschte und dokumentierte Umweltauswirkungen im Rahmen der LCA-Analyse für die Entscheidungsfindung berücksichtigt werden. Bislang unbekannt und zum jetzigen Zeitpunkt nicht ersichtliche Auswirkungen werden aufgrund mangelnder Kenntnis ausgeklammert, was der Methodik lediglich eine beratende und keinesfalls definitive Rolle in der Entscheidungsfindung zukommen lässt (Myllyviita et al. (2012), S. 245).

Hinsichtlich der Wahl geeigneter MADM-Methoden für dieses Anwendungsgebiet zeigen sich sowohl die Ansätze der amerikanischen als auch die der europäischen Schule als geeignet. Vorteil für erstere ist vor allem die praktischere Handhabung. Fraglich ist jedoch, ob die aufgrund des hohen Aggregationsgrades mögliche Kompensation von negativen und positiven Kriterienausprägungen dieser Ansätze für das Treffen von nachhaltigen Entscheidungen erwünscht ist. Besonders in dieser Hinsicht zeigen sich die Outranking-Ansätze der europäischen Schule als deutlich weniger anfällig (Ober-schmidt et al. (2010), S. 185 f.; Cinelli et al. (2013)).

Insgesamt erscheint die Kombination von MCDA und LCA aufgrund der sich gegenseitig ergänzenden Stärken jedoch sinnvoll. Die Lebenszyklusanalyse kann durch die in Kapitel 4.2 beschriebene Datenerfassung und -verarbeitung möglicher Risiken und Unsicherheiten entscheidend dabei helfen, eine dynamische Komponente im Entscheidungsprozess methodisch zu erfassen. Die Wahl einer geeigneten MCDA-Methode unterstützt darüber hinaus Entscheidungsträgern dabei, diese generierten Daten sinnvoll im Entscheidungsprozess zu nutzen (Scott et al. (2016), S. 1918).

6.1.3. Methodische Entwicklung im Forschungsfeld der DMCD-Modelle

Der in Kapitel 4.3 präsentierte Forschungsstrang zu DMCD-Modellen berücksichtigt eine zeitliche Komponente auf zwei verschiedenen Wegen. Zum einen werden vergangene Informationen über Alternativen und Kriterien in einem iterativen Prozess in die gegenwärtige Entscheidungsfindung einbezogen und somit eine Lernfähigkeit modelliert. Zum anderen wird ebenso eine zukunftsbezogene Komponente durch die Prognose einer zukünftigen Entscheidungsmatrix inkludiert. Ebenso besteht somit die Möglichkeit, Kriterien und Alternativen hinzuzufügen oder zu entfernen (Jassbi et al. (2014), S. 236). Diese Feedforward-Funktion ermöglicht es dem Entscheidungsträger, ein zukünftiges Szenario ohne weitere methodische Ergänzungen zu berücksichtigen. Gänzlich ersetzen können die präsentierten DMCD-Modelle den kombinierten Einsatz von MCDA und Szenarioplanung jedoch noch nicht, da jeweils nur eine globale Gesamt-aggregation über alle Entscheidungsmatrizen durchgeführt wird und für eine vergleichende Auswertung mehrerer Szenarien somit mehrere Modelle parallel aufgestellt werden müssten.

Abhilfe hierzu könnten jedoch die von Zolfani et al. vorgeschlagenen PMADM-Ansätze schaffen, welche diese zeitliche Komponente über Limitierungen bzw. „negative Szenarien“ versuchen zu berücksichtigen. Anwendungsbeispiele für diesen Forschungsansatz ließen sich im Rahmen des Suchprozesses dieser Arbeit jedoch nicht identifizieren.

Ein Schwachpunkt der in Kapitel 4.3 präsentierten Methode von Campanella und Ribeiro ist sicherlich die Tatsache, dass sie nicht für Konsensentscheidungen ausgelegt ist, sondern lediglich für Anwendung durch einen einzelnen Entscheidungsträger geeignet ist (Campanella und Ribeiro (2011), S. 56). Dies gestaltet eine praktische Anwendung für strategische Entscheidungen schwierig, handelt es sich hierbei doch vorrangig um in der Gruppe getroffene Entscheidungen.

6.2. Auffälligkeiten und weiterer Forschungsbedarf

Die Ergebnisse der Suche offenbarten verschiedene Methoden zum Umgang mit Dynamik im Entscheidungsprozess und zahlreiche Anwendungsmöglichkeiten für diese. Das vielschichtige und variierte Verständnis von Dynamik in der Literatur und Forschung im Zusammenhang mit multi-kriteriellen Entscheidungsproblemen erschwerte den Suchprozess in den in Kapitel 3.2 genannten Datenbanken. Viele Autoren verwendeten nicht explizit den Begriff der Dynamik in den Titeln und Schlagworten Ihrer Arbeit. Zu den in Kapitel 2.2 vorgestellten MODM- und MADM-Methoden existieren bereits zahlreiche Reviewartikel zu unterschiedlichen Fragestellungen und Problemkontexten. So untersuchen Mendoza und Martins (2006) in Ihrer Arbeit die Anwendbarkeit von MCDA auf dem Gebiet des Managements natürlicher Ressourcen, Behzadian et al. (2010) untersuchen 217 Arbeiten aus wissenschaftlichen Zeitschriften auf PROMETHEE-Methoden und Ihre Anwendungsgebiete und Mardani et al. (2015a) behandeln extensiv die Anwendung von MCDA-Methoden und speziell Fuzzy MCDA im Zeitraum von 1994 bis 2014. Bei Scott et al. (2012) findet sich ein Reviewbeitrag zum Einsatz von MCDA-Methoden im Kontext von Bionergiesystementscheidungen und Ho et al. (2010) prüfen verwendete MCDA-Methoden bei der Lieferantenauswahl auf verwendete Kriterien und mögliche Verbesserungen. Weniger bestehende Reviewbeiträge wurden hingegen zu dynamischen MCDA-Methoden identifiziert. Eine aktuelle Arbeit zu ebendiesen von Zolfani et al. (2016a) stellt im Rahmen eines Reviews die Eignung von MADM-Methoden für den dynamischen Problemkontext heraus, diskutiert jüngste Beiträge und expliziert weiteren Forschungsbedarf auf diesem Gebiet.

Betrachtet man die als relevant klassifizierten Beiträge in Tabelle 2, fällt unter anderem die verhältnismäßig hohe Anzahl an Artikeln aus Journals ohne Ranking im Feld der DMCD-Ansätze auf. Liegt die Quote der in Journals mit einem A-Ranking veröffentlichten Beiträge zu Szenario-basierten Ansätzen noch bei 32%, so sind lediglich 20% der Beiträge zum Thema der DMCD in einem Journal mit Ranking von A oder B veröffentlicht. Der überwiegende Anteil (53%) der Beiträge zu diesem Themengebiet ist hierbei in

vom Verband für Hochschullehrer e.V. nicht bewerteten Journals erschienen. Die Quote der in A oder B-Journals veröffentlichten Artikel zu MCDA und Lebenszyklusanalyse beträgt 50%. Da im Rahmen dieser Arbeit ausdrücklich kein Anspruch auf eine vollständige Erfassung sämtlicher relevanter Literatur gelegt wurde bleibt offen, ob diese Auffälligkeit auf der gewählten Charakteristik dieses Reviewbeitrages oder einer Forschungslücke im vergleichsweise jungen Gebiet der DMCD-Methoden beruht. Für letztere These spricht zudem die hohe Anzahl an Anwendungsbeispielen innerhalb der Arbeiten zu den in Kapitel 4.1 und 4.2 präsentierten Methoden, während die repräsentativ für das Forschungsfeld der DMCD-Modelle betrachteten Ansätze eher methodische Rahmenkonzepte besprechen und in numerischen Beispielen anstelle praxisbezogener Anwendungsfälle angewandt werden.

7. Schlussbetrachtung

“Some reviews end abruptly with a short conclusion. However, even though you have completed the majority of your review paper at this point, you can still tell your colleagues more. For instance, [...] to draw implications for practice and future theorizing” (Webster und Watson (2002), S. 20)

Existieren zu den etablierten Methoden der multikriteriellen Entscheidungsunterstützung bereits zahlreiche Anwendungsbeispiele sowie Reviewbeiträge, war es die Aufgabe dieser Arbeit explizit den derzeitigen State-of-the-Art des vergleichsweise jungen Forschungsfeldes der dynamischen MCDA-Methoden herauszustellen. Hierzu sollten, mithilfe einer strukturierten Literaturrecherche, Ansätze zur Integration einer zeitlichen Komponente in MCDA-Methoden herausgearbeitet werden.

Hierzu wurden 51 Beiträge aus wissenschaftlichen Zeitschriften sowie ausgewählten Buchveröffentlichungen nach einem methodischen Schema strukturiert ausgewertet und aufbereitet. Durch Anwendung dieses methodischen Schemas offenbarten sich drei grundsätzliche Ansätze zur Integration zeitlicher Aspekte in die MCDA, abhängig vom Problemkontext und dem Verständnis der Dynamik: Über die Kombination von MCDA-Methoden mit der Szenarioplanung, der Lebenszyklusanalyse oder über eigenständige DMCD-Modelle ohne weitere Hilfsmethodiken.

Die Analyse der Szenario-basierten Ansätze zeigte neben Rahmenkonzepten zudem einige Anwendungsbeispiele, was für eine fortgeschrittene Entwicklung in diesem Forschungsfeld spricht. Das vorgestellte Rahmenkonzept von Goodwin und Wright auf Basis der MAVT erfreut sich in diesem Forschungsfeld einer großen Beliebtheit. Als Anknüpfungspunkt für zukünftige Forschung wurde vorwiegend die Konstruktionsweise sowie die optimale Anzahl der verwendeten Szenarien identifiziert. Die Kombination von MCDA und LCA offenbart ihre Stärken vorallem im Anwendungsgebiet der nachhaltigen Entwicklung. Die Recherche offenbarte, dass auf die-

sem Gebiet bereits zahlreiche Anwendungsbeispiele vorliegen. Für eine hohe Ergebnisqualität eignet sich besonders die Verwendung von Outranking-Ansätzen, allerdings auf Kosten der Benutzerfreundlichkeit. Zudem erscheint es fraglich, ob eine vollständige Erfassung der zeitlichen Komponente im Sinne der vorliegenden Definition von Dynamik gewährleistet ist.

Als besser geeignet zu dieser vollständigen Erfassung erweisen sich die DMCD-Ansätze mit gleichzeitiger Berücksichtigung vergangener und zukünftiger Daten über mehrere Entscheidungsmatrizen. Dieser neuerliche Forschungsstrang birgt zudem noch am meisten Entwicklungspotential. Konkrete Ansatzpunkte hierzu sind etwa die Entwicklung einer Methodik für Konsensentscheidungen sowie das Untersuchen des Aggregationsprozesses.

Literatur

- Aouni, B., Colapinto, C., La Torre, D., Liuzzi, D., und Marsiglio, S. On dynamic multiple criteria decision making models: A goal programming approach. In *Multiple criteria decision making in finance, insurance and investment*, pages 31–48. Springer, 2015.
- Behzadian, M., Kazemzadeh, R. B., Albadvi, A., und Aghdasi, M. Promethee: A comprehensive literature review on methodologies and applications. *European journal of Operational research*, 200(1):198–215, 2010.
- Belton, V. und Stewart, T. *Multiple criteria decision analysis: an integrated approach*. Springer Science & Business Media, 2002.
- Belton, V. und Stewart, T. Problem structuring and multiple criteria decision analysis. In *Trends in multiple criteria decision analysis*, pages 209–239. Springer, 2010.
- Benítez, J., Delgado-Galván, X., Izquierdo, J., und Pérez-García, R. An approach to ahp decision in a dynamic context. *Decision Support Systems*, 53(3):499–506, 2012.
- Braune, I., Pinkwart, A., und Reeg, M. Book of abstracts / 5th Dubrovnik Conference on Sustainable Development of Energy, Water and Environment Systems, September 29th - October 3rd 2009, Dubrovnik, Croatia: UNESCO sponsored conference. Faculty of Mechanical Engineering and Naval Architecture, Zagreb 2009, 2009.
- Cairns, G., Goodwin, P., und Wright, G. A decision-analysis-based framework for analysing stakeholder behaviour in scenario planning. *European Journal of Operational Research*, 249(3):1050–1062, 2016.
- Campanella, G. und Ribeiro, R. A. A framework for dynamic multiple-criteria decision making. *Decision Support Systems*, 52(1):52–60, 2011.
- Campanella, G., Pereira, A., Ribeiro, R. A., und Varela, M. L. R. Collaborative dynamic decision making: A case study from b2b supplier selection. In *Euro Working Group Workshop on Decision Support Systems*, pages 88–102. Springer, 2011.
- Chen, Y., Li, K. W., und He, S. Dynamic multiple criteria decision analysis with application in emergency management assessment. In *Systems Man and Cybernetics (SMC), 2010 IEEE International Conference on*, pages 3513–3517. IEEE, 2010.
- Cinelli, M., Coles, R., und Kirwan, K. Use of multi criteria decision analysis to support life cycle sustainability assessment: An analysis of the appropriateness of the available methods. In *The 6th International Conference on Life Cycle Management*, pages 25–28, 2013.
- Cooper, H. M. Organizing knowledge syntheses: A taxonomy of literature reviews. *Knowledge in society*, 1(1):104–126, 1988.
- Cooper, H. und Hedges, L. *The Handbook of research synthesis*. Russell Sage Foundation, New York, 1994.
- Corner, J., Buchanan, J., und Henig, M. Dynamic decision problem structuring. *Journal of Multi-Criteria Decision Analysis*, 10(3):129–141, 2001.
- Crousillat, E. O., Dorfner, P., Alvarado, P., und Merrill, H. M. Conflicting objectives and risk in power system planning. *IEEE Transactions on Power Systems*, 8(3):887–893, 1993.
- Diakoulaki, D. und Grafakos, S. Externalities of Energy: Extension of Accounting Framework and Policy Applications. National Technical University Athens, 2004.
- Domingues, A. R., Marques, P., Garcia, R., Freire, F., und Dias, L. C. Applying multi-criteria decision analysis to the life-cycle assessment of vehicles. *Journal of cleaner production*, 107:749–759, 2015.
- Durbach, I. N. und Stewart, T. J. Modeling uncertainty in multi-criteria decision analysis. *European Journal of Operational Research*, 223(1):1–14, 2012.
- Ehrgott, M., Figueira, J., und Greco, S. *Trends in multiple criteria decision analysis*. Springer, New York, 2010.
- Fettke, P. State-of-the-art des state-of-the-art. *Wirtschaftsinformatik*, 48(4): 257, 2006.
- Figueira, J., Greco, S., Ehrgott, M., und Henggeler Antunes, C. Multiple criteria decision analysis: state of the art surveys. (International series in operations research & management science 78); Springer, New York.
- Fishburn, P. C. und Lavalley, I. H. Mcda: theory, practice and the future. *Journal of Multi-Criteria Decision Analysis*, 8(1):1–2, 1999.
- Frini, A. A multi-criteria multi-period approach for energy project selection in sustainable development context. In *Green Energy, 2014 International Conference on*, pages 65–71. IEEE, 2014.
- Gardfield, E. Reviewing review literature. part 1: definitions and uses of reviews. *Current Contents. Essays of an Information Scientist*, 10(18):113–116, 1987.
- Geldermann, J. und Rentz, O. Multi-criteria analysis for technique assessment: Case study from industrial coating. *Journal of Industrial Ecology*, 9(3):127–142, 2005.
- Gerking, H. Modeling of multi-stage decision-making processes in multi-period energy-models. *European Journal of Operational Research*, 32(2): 191–204, 1987.
- Goodwin, P. und Wright, G. Enhancing strategy evaluation in scenario planning: a role for decision analysis. *Journal of management studies*, 38(1): 1–16, 2001.
- Goodwin, P. und Wright, G. *Decision Analysis for Management Judgment 5th ed.* Number 5th. John Wiley and sons, 2014.
- Gorenstin, B., Campodonico, N., Costa, J., und Pereira, M. Power system expansion planning under uncertainty. *IEEE Transactions on Power Systems*, 8(1):129–136, 1993.
- Great Britain, Department for Communities and Local Government. Multi-criteria analysis: a manual. Communities and Local Government, Wetherby, 2009.
- Guinée, J. B. Handbook on life cycle assessment operational guide to the iso standards. *The international journal of life cycle assessment*, 7(5):311–313, 2002.
- Halog, A. und Manik, Y. Advancing integrated systems modelling framework for life cycle sustainability assessment. *Sustainability*, 3(2):469–499, 2011.
- Henig, M. I. und Buchanan, J. T. Solving mcdm problems: Process concepts. *Journal of Multi-Criteria Decision Analysis*, 5(1):3–21, 1996.
- Hermann, B., Kroeze, C., und Jawjit, W. Assessing environmental performance by combining life cycle assessment, multi-criteria analysis and environmental performance indicators. *Journal of Cleaner Production*, 15(18):1787–1796, 2007.
- Hites, R., De Smet, Y., Risse, N., Salazar-Neumann, M., und Vincke, P. About the applicability of mcdm to some robustness problems. *European Journal of Operational Research*, 174(1):322–332, 2006.
- Ho, W., Xu, X., und Dey, P. K. Multi-criteria decision making approaches for supplier evaluation and selection: A literature review. *European Journal of operational research*, 202(1):16–24, 2010.
- Jasch, C. Environmental performance evaluation and indicators. *Journal of Cleaner Production*, 8(1):79–88, 2000.
- Jassbi, J. J., Ribeiro, R. A., und Varela, L. R. Dynamic mcdm with future knowledge for supplier selection. *Journal of Decision Systems*, 23(3):232–248, 2014.
- Jiang, L., Liu, H., Martínez, L., und Cai, J. A model for linguistic dynamic multi-criteria decision-making. In *Foundations of Intelligent Systems*, pages 939–949. Springer, 2014.
- Kallrath, J. *Gemischt-ganzzahlige Optimierung: Modellierung in der Praxis: Mit Fallstudien aus Chemie, Energiewirtschaft, Metallgewerbe, Produktion und Logistik*. Springer-Verlag, 2013.
- Keynes, J. M. The general theory of employment. *The quarterly journal of economics*, 51(2):209–223, 1937.
- Kilgour, D. M., Chen, Y., und Hipel, K. W. Multiple criteria approaches to group decision and negotiation. In *Trends in multiple criteria decision analysis*, pages 317–338. Springer, 2010.
- Kluczek, A. Application of multi-criteria approach for sustainability assessment of manufacturing processes. *Management and Production Engineering Review*, 7(3):62–78, 2016.
- Kornbluth, J. Dynamic multi-criteria decision making. *Journal of Multi-Criteria Decision Analysis*, 1(2):81–92, 1992.
- Labuschagne, C. und Brent, A. C. Sustainable project life cycle management: the need to integrate life cycles in the manufacturing sector. *International Journal of Project Management*, 23(2):159–168, 2005.
- Lienert, J., Scholten, L., Egger, C., und Maurer, M. Structured decision-making for sustainable water infrastructure planning and four future scenarios. *EURO Journal on Decision Processes*, 3(1-2):107–140, 2015.
- Light, R. und Pillemer, D. *Summing up*. Harvard University Press, 1984.
- Lin, Y.-H., Lee, P.-C., und Ting, H.-I. Dynamic multi-attribute decision making model with grey number evaluations. *Expert Systems with Applications*, 35(4):1638–1644, 2008.
- Linares, P. Multiple criteria decision making and risk analysis as risk management tools for power systems planning. *IEEE Transactions on Power Systems*, 17(3):895–900, 2002.
- Macombe, C., Leskinen, P., Feschet, P., und Antikainen, R. Social life cycle assessment of biodiesel production at three levels: a literature review and

- development needs. *Journal of Cleaner Production*, 52:205–216, 2013.
- Manten, A. A. Scientific literature review. *Scholarly Publishing*, 5:75–89, 1973.
- Mardani, A., Jusoh, A., MD Nor, K., Khalifah, Z., Zakwan, N., und Valipour, A. Multiple criteria decision-making techniques and their applications—a review of the literature from 2000 to 2014. *Economic Research-Ekonomska Istraživanja*, 28(1):516–571, 2015a.
- Mardani, A., Jusoh, A., und Zavadskas, E. K. Fuzzy multiple criteria decision-making techniques and applications—two decades review from 1994 to 2014. *Expert Systems with Applications*, 42(8):4126–4148, 2015b.
- Matsatsinis, N. F. und Samaras, A. P. Mcdm and preference disaggregation in group decision support systems. *European Journal of Operational Research*, 130(2):414–429, 2001.
- Mendoza, G. A. und Martins, H. Multi-criteria decision analysis in natural resource management: a critical review of methods and new modelling paradigms. *Forest ecology and management*, 230(1-3):1–22, 2006.
- Millán, J., Campo, R., und Sánchez-Sierra, G. A modular system for decision-making support in generation expansion planning (super). *IEEE transactions on power systems*, 13(2):667–671, 1998.
- Miranda, V. und Proenca, L. M. Why risk analysis outperforms probabilistic choice as the effective decision support paradigm for power system planning. *IEEE Transactions on Power Systems*, 13(2):643–648, 1998.
- Montibeller, G., Gummer, H., und Tumidei, D. Combining scenario planning and multi-criteria decision analysis in practice. *Journal of Multi-Criteria Decision Analysis*, 14(1-3):5–20, 2006.
- Munda, G. Multiple criteria decision analysis and sustainable development. In *Multiple Criteria Decision Analysis*, pages 1235–1267. Springer, 2016.
- Münzberg, T., Wiens, M., und Schultmann, F. Dynamic-spatial vulnerability assessments: a methodical review for decision support in emergency planning for power outages. *Procedia Engineering*, 78:78–87, 2014.
- Myllyviita, T., Holma, A., Antikainen, R., Lähtinen, K., und Leskinen, P. Assessing environmental impacts of biomass production chains—application of life cycle assessment (lca) and multi-criteria decision analysis (mcdm). *Journal of cleaner production*, 29:238–245, 2012.
- Oberschmidt, J., Geldermann, J., Ludwig, J., und Schmehl, M. Modified promethee approach for assessing energy technologies. *International Journal of Energy Sector Management*, 4(2):183–212, 2010.
- Omam, I. *Multi-criteria decision aid as an approach for sustainable development analysis and implementation*. Graz, 2004.
- Onat, N. C., Kucukvar, M., Tatari, O., und Zheng, Q. P. Combined application of multi-criteria optimization and life-cycle sustainability assessment for optimal distribution of alternative passenger cars in us. *Journal of Cleaner Production*, 112:291–307, 2016.
- Orji, I. J. und Wei, S. Dynamic modeling of sustainable operation in green manufacturing environment. *Journal of Manufacturing Technology Management*, 26(8):1201–1217, 2015.
- Pais, T. C. und Ribeiro, R. A. Contributions to dynamic multicriteria decision making models. In *IFSA/EUSFLAT Conf*, pages 719–724. Citeseer, 2009.
- Phillips, L. *Tackling strategic problems: the role of group decision support*. SAGE Publications, London; Newbury Park, Calif, 1990.
- Pictet, J. und Bollinger, D. The silent negotiation or how to elicit collective information for group mcdm without excessive discussion. *Journal of Multi-Criteria Decision Analysis*, 13(5-6):199–211, 2005.
- Polatidis, H., Haralambopoulos, D. A., Kemp, R., und Rothman, D. Creating an energy system that we want but don't know yet, using integrated assessment, transition management and multi-criteria analysis. *Integrated Assessment*, 4(3):205–213, 2003.
- Polatidis, H., Haralambopoulos, D. A., Munda, G., und Vreeker, R. Selecting an appropriate multi-criteria decision analysis technique for renewable energy planning. *Energy Sources, Part B*, 1(2):181–193, 2006.
- Ram, C. und Montibeller, G. Exploring the impact of evaluating strategic options in a scenario-based multi-criteria framework. *Technological Forecasting and Social Change*, 80(4):657–672, 2013.
- Reichert, P., Langhans, S. D., Lienert, J., und Schuwirth, N. The conceptual foundation of environmental decision support. *Journal of environmental management*, 154:316–332, 2015.
- Rezaei, J. Best-worst multi-criteria decision-making method. *Omega*, 53:49–57, 2015.
- Ribeiro, F., Ferreira, P., und Araújo, M. Evaluating future scenarios for the power generation sector using a multi-criteria decision analysis (mcdm) tool: The portuguese case. *Energy*, 52:126–136, 2013.
- Saaty, T. L. Time dependent decision-making; dynamic priorities in the ahp/analytic hierarchy process: Generalizing from points to functions and from real to complex variables. *Mathematical and Computer Modelling*, 46(7-8):860–891, 2007.
- Schmuck, P. et al. Transdisciplinary evaluation of energy scenarios for a german village using multi-criteria decision analysis. *Sustainability*, 4(4):604–629, 2012.
- Scholten, L., Scheidegger, A., Reichert, P., Mauer, M., und Lienert, J. Strategic rehabilitation planning of piped water networks using multi-criteria decision analysis. *Water research*, 49:124–143, 2014.
- Scholten, L., Schuwirth, N., Reichert, P., und Lienert, J. Tackling uncertainty in multi-criteria decision analysis—an application to water supply infrastructure planning. *European Journal of Operational Research*, 242(1):243–260, 2015.
- Scott, J. A., Ho, W., und Dey, P. K. A review of multi-criteria decision-making methods for bioenergy systems. *Energy*, 42(1):146–156, 2012.
- Scott, R. P., Cullen, A. C., Fox-Lent, C., und Linkov, I. Can carbon nanomaterials improve czts photovoltaic devices? evaluation of performance and impacts using integrated life-cycle assessment and decision analysis. *Risk Analysis*, 36(10):1916–1935, 2016.
- Stewart, T. J. A critical survey on the status of multiple criteria decision making theory and practice. *Omega*, 20(5-6):569–586, 1992.
- Stewart, T. J., French, S., und Rios, J. Integrating multicriteria decision analysis and scenario planning—review and extension. *Omega*, 41(4):679–688, 2013.
- Stewart, T. und Durbach, I. Dealing with Uncertainties in MCDA; in: *Multiple Criteria Decision Analysis*, Band 233; hrsg. v. Greco S, Ehrgott M, Figueira JR; Springer New York, S. 467–496, 2016.
- Vaillancourt, K. und Waub, J.-P. Equity in international greenhouse gases abatement scenarios: A multicriteria approach. *European Journal of Operational Research*, 153(2):489–505, 2004.
- Virgo, J. A. The review article: its characteristics and problems. *The Library Quarterly*, 41(4):275–291, 1971.
- Watrobski, J., Jankowski, J., und Ziemia, P. Multistage performance modelling in digital marketing management. *Economics & Sociology*, 9(2):101–125, 2016.
- Webster, J. und Watson, R. T. Analyzing the past to prepare for the future: Writing a literature review. *MIS quarterly*, pages xiii–xxiii, 2002.
- Wilkens, I. *Multikriterielle analyse zur nachhaltigkeitsbewertung von energiesystemen-von der theorie zur praktischen anwendung*. 2012.
- Wright, G. und Goodwin, P. Future-focussed thinking: combining scenario planning with decision analysis. *Journal of Multi-Criteria Decision Analysis*, 8(6):311–321, 1999.
- Xu, B. und Ouenniche, J. Performance evaluation of competing forecasting models: A multidimensional framework based on mcdm. *Expert Systems with Applications*, 39(9):8312–8324, 2012.
- Yu, P.-L. und Chen, Y.-C. Dynamic mcdm, habitual domains and competence set analysis for effective decision making in changeable spaces. In *Trends in multiple criteria decision analysis*, pages 1–35. Springer, 2010.
- Zahir, S. Aggregation of priorities in multi-criteria decision analysis (mcdm): Connecting decision spaces in the cognitive space. *American Journal of Operations Research*, 6(04):317–333, 2016.
- Zheng, J., Egger, C., und Lienert, J. A scenario-based mcdm framework for wastewater infrastructure planning under uncertainty. *Journal of environmental management*, 183:895–908, 2016.
- Zimmermann, H.-J. und Gutsche, L. *Multi-Criteria Analyse*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1991.
- Zolfani, S., Maknoon, R., und Zavadskas, E. K. An introduction to prospective multiple attribute decision making (pmadm). *Technological and Economic Development of Economy*, 22(2):309–326, 2016a.
- Zolfani, S. H., Maknoon, R., und Zavadskas, E. K. Madm and futures studies; a necessity. In *Conference: 9th International Scientific Conference "Business and Management"*, 2016b.
- Zulueta, Y., Martínez-Moreno, J., Martínez, L., und Espinilla, M. A discriminative dynamic index based on bipolar aggregation operators for supporting dynamic multi-criteria decision making. In *Aggregation Functions in Theory and in Practice*, pages 237–248. Springer, 2013.