**Master's Degree Program at FHWien der WKW in Organizational & Human Resources Development**

# EVALUATION OF MULTI-STAKEHOLDER INITIATIVES – THE SUITABILITY OF CURRENT EVALUATION TOOLS FOR MULTI-STAKEHOLDER INITIATIVES IN GLOBAL CLIMATE GOVERNANCE

**Desired Academic Title:**
**Master of Arts in Business**

**Submitted by: Karoline Kehrer**
**Matriculation Number: 00440585**
**Class of: 2022**
**Supervisor: Maria Riegler**

I hereby confirm that I,

- Wrote this work on my own, did not use any other sources and aids than the ones mentioned and did not get help in any other way,
- Have not yet submitted this work in any form as an examination paper either at home or abroad,
- Am submitting the version of this thesis that was submitted to the supervisor and uploaded for plagiarism checking,
- Agree to the publication of this work by the library at FHWien der WKW, which will take place after the expiry of the approved period in the case of a block.

___Vienna, 25.08.2022_____

Place, Date Signature

## Abstract

Multi-stakeholder initiatives (MSIs) are attributed a major role in resolving global social and environmental problems, yet their legitimacy and effectiveness are disputed. The evaluation of MSIs is essential to ensure their quality. Research knowledge in this field, however, is limited. This master thesis contributes to improving MSI evaluation by analyzing current evaluation tools with regard to their suitability for MSIs in climate governance, a global concern of particular urgency. For empirical research, I used document analysis as a method of data collection, and qualitative content analysis according to Mayring as a method of data analysis. The results show that while current evaluation tools cover a variety of aspects, they mirror the deficits of MSIs criticized by researchers and are scarcely equipped for assessing power imbalances, accountability mechanisms, and the impact of MSIs. Also, particular challenges and requirements of climate-related MSIs are not sufficiently represented. Although existing tools can be a good starting point for evaluation, they require research-based improvement, including the development of sector-specific tools for the evaluation of MSIs in global climate governance.

Multi-Stakeholder-Initiativen (MSIs) wird eine wichtige Rolle bei der Lösung globaler sozialer und ökologischer Probleme zugeschrieben, ihre Legitimität und Wirksamkeit sind jedoch umstritten. Die Evaluierung von MSIs ist unerlässlich, um ihre Qualität zu gewährleisten. Das Forschungswissen in diesem Bereich ist allerdings begrenzt. Diese Masterarbeit leistet einen Beitrag zur Verbesserung der Evaluierung von MSIs, indem sie bestehende Evaluierungsinstrumente auf ihre Eignung für MSIs im Klimaschutz, einem globalen Anliegen von besonderer Dringlichkeit, untersucht. Die empirische Forschung führte ich mittels Dokumentenanalyse als Methode der Datenerhebung und qualitativer Inhaltsanalyse nach Mayring als Methode der Datenanalyse durch. Die Ergebnisse zeigen, dass die derzeitigen Evaluierungsinstrumente zwar eine Vielzahl von Aspekten abdecken, jedoch die von Forschenden kritisierten Defizite von MSIs widerspiegeln und kaum geeignet sind, Machtungleichgewichte, Rechenschaftsmechanismen und den ‚Impact' von MSIs zu bewerten. Auch die besonderen Herausforderungen und Ansprüche klimabezogener MSIs werden nicht ausreichend abgebildet. Obwohl die bestehenden Instrumente einen guten Ausgangspunkt für die Evaluierung darstellen können, ist eine forschungsbasierte Weiterentwicklung notwendig, einschließlich der Erarbeitung von spezifischen Instrumenten für die Evaluierung von MSIs im Klimabereich.

## Acknowledgements

I would like to express my appreciation to all who supported me in accomplishing this master thesis. My thanks go to my supervisor, Maria Riegler, for her valuable feedback, which always was spot on and helped me improve my work. I am also grateful to everyone who helped me enhance this thesis by proofreading for their valuable time and support. I particularly want to thank Martina for her dedication and thoroughness, and Lisa, Konstantin, and Jay for taking some of the last-minute stress off me.

Furthermore, I appreciate the numerous productive writing sessions with Fabian, which never lacked some fun, either. Finally, I want to thank my colleagues Juliane and Eva for the continuous great teamwork, exchange, and mutual support throughout the entire master's program.

## Table of Contents

## List of Abbreviations

| | |
|---|---|
| CEO | Chief Executive Officer |
| CSO | Chief Security Officer |
| CSR | Corporate Social Responsibility |
| cf. | confer |
| DOCX | document extended (referring to Microsoft Word 2007/2010/2013 documents) |
| etc. | et cetera |
| EU | European Union |
| GHG | greenhouse gas |
| GRI | Global Reporting Initiative |
| HR | Human Resources |
| ID | Identification |
| i.e. | that is (the abbreviation comes from Latin ‘id est’, meaning ‘that is’ in English) |
| IGO | Intergovernmental Organization |
| MSIs | Multi-stakeholder Initiatives |
| NGO | Non-Governmental Organization |
| no. | number |
| OECD | Organisation for Economic Co-operation and Development |
| OECD DAC | Organisation for Economic Co-operation and Development Development Assistance Committee |
| SD | Sustainable Development |
| SDGs | Sustainable Development Goals |
| UN | United Nations |
| UNFCCC | United Nations Framework Convention on Climate Change |
| UN GC | United Nations Global Compact |
| UNGPs | United Nations Guiding Principles on Business and Human Rights |
| US | United States (of America) |

## List of Figures

## List of Tables

# 1 Introduction

## 1.1 Problem Statement

Multi-stakeholder initiatives (MSIs) are collaborative, cross-sectoral efforts between various actors that target current social or environmental challenges. In literature, such collective action efforts can also be found under different terms (see Table 1) which all, with slightly varying definitions and scope, describe similar collaborations (e.g. Global Development Incubator, 2015; Jastram & Klingenberg, 2018; Mena & Palazzo, 2012).

Multi-stakeholder initiatives have grown in number as well as in importance and influence since the 1990s in the context of the 1992 Earth Summit and later conferences on sustainable development, up to the establishment of the Sustainable Development Goals by the United Nations in 2015. They have been introduced as an instrument to fill governance gaps not covered by traditional nation-state governments and international agreements in the complex area of sustainable development, and are today considered significant contributors to global governance (e.g. Eweje et al., 2021; Oguntuase, 2020; Pattberg & Widerberg, 2016). In Sustainable Development Goal 17: "Strengthen the means of implementation and revitalize the Global Partnership for Sustainable Development", the 2030 Agenda for Sustainable Development of the United Nations Agenda explicitly acknowledges the contribution of multi-stakeholder partnerships and commits to their enhancement (United Nations, 2015b). Global climate governance with its complex and global dynamics and increasing urgency particularly requires the involvement of a broad range of actors across different sectors (Andonova et al., 2009).

MSIs face a range of expectations and demands regarding their mode and results of operation. Main concerns are their legitimacy, which includes participation, deliberative practice, transparency and accountability, and their effectiveness (Widerberg & Pattberg, 2015), both of which have increasingly been questioned by critics (e.g. Bäckstrand, 2012; Biermann et al., 2007; MSI Integrity, 2020a; Pattberg & Widerberg, 2016). While MSIs are considered important contributors to solving global social and environmental problems, little evidence exists for their positive performance (Pattberg & Widerberg, 2016).

The findings of Pattberg and Widerberg (2016) and MSI Integrity and International Human Rights Clinic at Harvard Law School (2017) underline the necessity to evaluate multi-stakeholder initiatives to ensure their effectiveness and legitimacy. Upon an analysis of 340 multi-stakeholder partnerships, Pattberg and Widerberg (2016) found that 62% of them where either

inactive, or delivering either no outputs or outputs diverging from their function. Only 24% of them generated an output fully congruent with their aspirations. Furthermore, multi-stakeholder initiatives in many cases do not fill the present governance gaps but rather tend to perpetuate existing power balances and marginalization of stakeholders, and their aptitude as instrument to meet governance deficits is called into question (MSI Integrity, 2020a; Pattberg & Widerberg, 2016). Still, the knowledge and research about the evaluation of MSIs are limited and few studies have systematically investigated the evaluation of collective action efforts (e.g. Biermann et al., 2007; Stadtler, 2016; Ven et al., 2017). However, there is indication of the lack of evaluation within MSIs (OECD, 2008).

Several evaluation tools and frameworks have been developed to remedy this shortcoming and facilitate the evaluation of collaborative efforts, which vary in scope, target sector, level of detail, etc. (e.g. Afsana et al., 2009; Andonova et al., 2009; J. M. Brinkerhoff, 2002; Marriott & Goyder, 2009; MSI Integrity & International Human Rights Clinic at Harvard Law School, 2017; OECD, 2008). Some reviews exist on single tools (e.g. Chianca, 2008), but barely any systematic research assessing and comparing different evaluation tools can be found.

Against this background, the present master thesis aims to deepen the knowledge on MSI evaluation tools, by analyzing currently available tools with regard to their suitability for MSIs in global climate governance.

## 1.2 Research Questions

Based on the problem statement, in this master thesis I investigate the following research questions:

The main research question is: “How suitable are currently existing evaluation tools for evaluating multi-stakeholder initiatives in global climate governance?”

The first sub-question addresses, “What evaluation aspects do current evaluation tools cover and how does this compare to the scientific literature?”. The second sub-question investigates, “What limitations regarding their applicability can be identified, generally and concerning the field of global climate governance?”.

## 1.3 Research Objective

This master thesis aims to provide new insights to advance the evaluation of multi-stakeholder initiatives, specifically in the area of climate governance.

Although MSIs are supposed to play a key role in solving our world's social and environmental problems, critics increasingly call their legitimacy and effectiveness into question, and there is little evidence for their successful performance (cf. Bäckstrand, 2012; Biermann et al., 2007; MSI Integrity, 2020a; Pattberg & Widerberg, 2016). In order to ensure their legitimacy and effectiveness, it is crucial to evaluate multi-stakeholder initiatives (cf. MSI Integrity & International Human Rights Clinic at Harvard Law School, 2017; Pattberg & Widerberg, 2016). Although various tools and frameworks have been developed for that purpose, their adequacy has hardly been investigated.

The objective of my master thesis is to expand the knowledge on this topic by systematically analyzing existing evaluation tools and investigating their suitability for the evaluation of MSIs in the field of global climate governance. I chose to direct my investigation to MSIs operating in global climate governance because of its particularly multi-stakeholder character and its pressing urgency (Andonova et al., 2009; Jagers & Stripple, 2003).

In a first step, I explore which aspects of evaluation currently available tools already cover, and in a second step, I compare them to the current scientific knowledge. I further examine the analysis results in terms of possible limitations to the applicability of the tools in general and from the perspective of climate governance.

In summary, the objective of my master thesis is to contribute to the literature by means of a systematic analysis of current evaluation tools, providing new findings about the suitability and applicability of current evaluation tools for MSIs in the climate sector.

## 1.4 Structural Outline

In the first part of this work, I provide an overview of the theoretical background found in (scientific) literature as a basis for the research objective of this master thesis. At the beginning, I clarify basic terms and concepts which are underpinnings of the subsequent chapters and conceptualizations. I then give an outline of scientific literature on MSIs and the origin and development of multi-stakeholder initiatives during the past decades. In the following subchapters, I describe the main functions and requirements which MSIs are supposed to fulfill, including the two main areas of legitimacy and effectiveness. Next, I outline the role and status quo of evaluation of multi-stakeholder initiatives and briefly introduce some existing evaluation tools and frameworks. To conclude the theoretical background, I summarize the main points in relation to my research aim.

In the second part of this master thesis, I present my empirical research. To begin with, I specify the design and methods of my research, explaining the document analysis used for data collection and the qualitative content analysis applied for data analysis. I also elaborate on the actual execution of the research, including encountered challenges and recommendations. In the following chapter, I describe the results of the empirical analysis. First, I summarize the evaluation aspects covered by the analyzed evaluation tools and briefly depict the quantitative category frequencies of the analysis. I subsequently display further results of the analysis based on overlaps in the categorization process and regarding the specificity of the evaluation questionnaires towards target groups. In the next step, I discuss the empirical results in relation to the scientific background provided in the first part of this thesis and in relation to my research objective, concluding with an answer to my research questions. In the remaining two subchapters, I examine the implications of my findings for the domain of human resources and organizational development, and end with an outlook on future research possibilities. Finally, I provide a short summary of this master thesis. The literature references can be found at the end of this work.

# 2 Theoretical Background

This chapter serves as a basis for comprehending the research objective, its background, and the current state of research on multi-stakeholder initiatives and of evaluation practices. At the beginning, I will introduce the basic terms and concepts for this thesis. Then I will briefly delineate the state of scientific literature on multi-stakeholder initiatives. Subsequently, I will provide a summary of the emergence and development of MSIs over the past decades, which also addresses MSIs in global climate governance and the growing criticism of MSIs. The main functions and characteristics that are promised by and demanded of MSIs follow next, encompassing the two main claims of legitimacy and effectiveness in addressing global social and environmental problems. Finally, I will outline the evaluation of multi-stakeholder initiatives, including its role and status quo, and introduce some evaluation frameworks. Concluding this chapter, I will summarize the main points in the context of my research question.

## 2.1 Terminology

In this chapter I will outline the basic concept of collective action and the terms used in literature to describe collective action efforts, such as 'multi-stakeholder initiatives' and others, as a basis for a clear and congruent understanding of the master thesis. As will become apparent in this chapter, a variety of similar denominations is present in scientific literature; there is no universal definition and their use and scope differ according to the authors. The following section aims to give an overview of the terminology landscape and clarify the further use of terms in this master thesis.

### 2.1.1 Collective Action

To help understand what 'collective action' is and what its relevance and challenges are, this chapter summarizes the core concept of collective action and its connection to present-day societal and environmental problems.

Collective action occurs when two or more individuals coordinate their decisions and actions in an interdependent situation in which the contribution of each individual influences the joint outcome (Ostrom, 2010; Sandler, 2010). In this constellation, individuals who seek to achieve the maximum joint outcome in most cases will generate reduced individual benefit and vice versa (Ostrom, 2010, 2015).

The question hereby is whether individuals accept a semi-optimal equilibrium of their own benefit to enable an optimal collective outcome. Collective action is frequently associated with the use of public or common goods which are often non-excludable (e.g. climate change mitigation or air pollution removal). That means that a single person cannot be excluded from profiting from the benefit provided by others and entails the so-called free-rider problem: For each individual actor it is tempting to not contribute to the joint effort but still benefit from its results, thus free-riding on the contributions of others. If too many actors decide to free-ride, an optimal collective benefit cannot be achieved; if all actors choose to free-ride, the collective objective will not be reached at all (Ostrom, 2015; Sandler, 2010).

Academic research and debate about collective action problems and underlying social theories have been undertaken for decades and are ongoing (cf. e.g. Olson, 1965; Ostrom, 2010). More recently, scholars have also increasingly focused on how collective action can be used to contribute to sustainable development, how to engage people in collective action, and how collective action can be successfully coordinated (e.g. Louis, 2009; Ostrom, 2015, pp. 58–102). To elaborate in detail on these debates and findings would exceed the scope of this master thesis, therefore I will not discuss them in this work.

Collective action is essential for societies and its necessity further increases with incremental globalization and associated complexity and interdependency of societies (Sandler, 2010). Many of the current problems which societies face are related to public goods and are, moreover, often global phenomena with a world-wide interdependency, such as most environmental problems. When it comes for example to climate change, regulations and measures of national governments are not sufficient to solve the problem and cooperation of more comprehensive groups of actors are required (Weimann et al., 2019). The Global Development Incubator sums it up as follows: “In essence, the S[ustainable] D[evelopment] G[oal]s are collective action problems” (Global Development Incubator, 2015, p. 6).

### 2.1.2 Denomination of Collective Action Efforts

In scientific and management literature, collective action efforts are discussed under a range of names which vary in use, delimitation, and scope, and mostly lack universal definitions. Different terms might be used to describe the same subject and at the same time, one term might be used inconsistently across publications. Basically, each scientific publication (re-)defines

their own demarcation of the employed term(s) (e.g. Global Development Incubator, 2015; Jastram & Klingenberg, 2018). Mena and Palazzo (2012, p. 533) attribute this phenomenon to "the mushrooming of those private regulatory initiatives in the last decade".

To give an impression and overview of the variety of terms found in literature, Table 1 shows a (non-exhaustive) list of denominations used for collective action efforts in global governance, based on listings by other authors (Global Development Incubator, 2015; Jastram & Klingenberg, 2018; Mena & Palazzo, 2012; Rasche, 2012) and my own research.

| | | |
|---|---|---|
| Multi-stakeholder initiatives | Collective action initiatives | Multi-stakeholder partnerships |
| Public-private partnerships | Multi-sector partnerships | Global governance |
| Collective impact (initiatives) | Partnership networks | Global partnerships |
| Cross-sector alliances | Multi-stakeholder arrangements | Global action networks |
| Voluntary commitments | Sustainable development action networks | Multi-stakeholder governance |
| International cooperative initiatives | Transnational partnerships | Governance experiments |
| Tripartite partnerships | Cross-sector partnerships | Transnational institutions |
| Private-to-private partnerships | Regulatory initiatives | Community-based partnerships |
| Networked governance | Multi-stakeholder networks | Transnational private regulation |
| Transnational norm-building networks | International certifiable standards | Global public policy networks |
| Collaborative alliances | Solution networks | Hyper collaborative partnerships |

*Table 1*. Examples for denominations of collective action efforts in global development. (Based on listings by Global Development Incubator, 2015; Jastram & Klingenberg, 2018; Mena & Palazzo, 2012; Rasche, 2012; and my own research.)

### 2.1.3 Global Governance

Concerning the term and concept of global governance, various scholars refer to the work of James Rosenau and their distinction between the terms ‘government’ and ‘governance’ (Andonova et al., 2009, p. 55; Bulkeley et al., 2012, p. 593; Bulkeley & Schroeder, 2012, p. 745; Domínguez & Velázquez Flores, 2012, sp; Weiss & Wilkinson, 2014, p. 208). While government is understood in the context of nation-states, governance encompasses the coordination of nation-states and other actors and rule systems beyond national jurisdiction on a global scale (Andonova et al., 2009, p. 55; Bulkeley et al., 2012, p. 593). It presumes the existence of shared goals and interests beyond the nation-state; therefore collective goods and collective action problems take an essential place in the conceptualization of global governance (Domínguez & Velázquez Flores, 2012, sp; Risse, 2012, p. 700; Weiss & Wilkinson, 2014, p. 213; Zürn, 2012, pp. 730–731). Risse (2012, p. 700) for example defines governance as “the various institutionalized modes of social coordination to produce and implement collectively binding rules, or to provide collective goods”. Zürn (2012, p. 730) describes global governance as the entirety of policies and regulations, structures, and processes introduced “with reference to solving specific denationalized and deregionalized problems or providing transnational common goods”. However, Zürn (2012, p. 730) also cautions that even though governance activities are promoted based on the common good, they are not automatically beneficial to it.

### 2.1.4 Multi-Stakeholder Initiatives (MSIs)

‘Multi-stakeholder initiatives’ is a term frequently found in literature to describe collective action efforts in sustainable development (e.g. Arenas et al., 2020; Bakker et al., 2019; Biekart & Fowler, 2018; Jastram & Klingenberg, 2018; Mena & Palazzo, 2012; MSI Integrity, 2020a; Rasche, 2012).

As I have already mentioned with regard to collective action efforts in general, no universal definition exists for multi-stakeholder initiatives, either. Instead, the term is defined in each publication individually, leading to a range of slightly varying definitions across literature.
Rasche (2012) describes MSIs as “a collaborative form of governance for CSR issues voluntarily involving an array of stakeholders, which, as a whole, cross the state/non-state and profit/non-profit boundaries” (Rasche, 2012, pp. 682–683). Another, more comprehensive definition of MSIs has been formulated by the Global Development Incubator (2015):

> MSIs are [temporary] organizations (1) focused on bringing about collective action solutions for global public benefit, (2) comprised of actors across the public and private sectors (both for-profit and philanthropic), and (3) whose governance bodies and capabilities are wholly new, rather than simply reliant on those of the constituent actors. (Global Development Incubator, 2015, p. 7)

What many definitions of MSIs across literature include are (1) the purpose of a social and/or environmental benefit and (2) cross-sectoral features, which involve at least two types/groups of stakeholders (e.g. Bakker et al., 2019; Fransen, 2012; Global Development Incubator, 2015; Jastram & Klingenberg, 2018; Rasche, 2012), setting them apart from traditional industry-specific regulations and trade associations.

As a prerequisite, the Global Development Incubator (2015) adds the participation of more than two actors and does not include a simple public-private partnership of two actors in the scope of MSIs. Some authors, in addition to describing the number of actors and groups involved, go on to specify their share of involvement. For example, Bakker et al. (2019) do not consider initiatives with only faint multi-stakeholder features, such as non-corporate stakeholders holding exclusively advisory functions, and Fransen (2012, p. 166) establishes "governance structures allowing for an equal possibility of input among the different partners in steering the initiative" as a prerequisite. Regarding the time scope, the Global Development Incubator (2015) highlights the intentional choice of the word "initiative", pointing at the time- and goal-bound design and existence of MSIs.

For some authors (e.g. Mena & Palazzo, 2012; Rasche, 2012), 'multi-stakeholder initiatives' serves as an umbrella term for the wide range of different collective action efforts. In this master thesis, I will also use the expression 'multi-stakeholder initiatives' as a comprehensive term but will also employ other specific terms in accordance with their uses by the referenced authors.

### 2.1.5 Partnerships for Collective Action – Multi-Stakeholder Partnerships, Public-Private Partnerships, etc.

Collective action efforts named 'partnerships' are also widely present in literature, amongst others by the terms 'multi-stakeholder partnerships' or 'public-private partnerships' (see Table 1), which are used and put forward by the United Nations as well as numerous scholars (e.g. Bäckstrand, 2006, 2012; Biermann et al., 2007; Clarke & MacDonald, 2019; Liu et al., 2018; Oguntuase, 2020; United Nations, 2015a, 2018; Westman & Broto, 2018).

Some other authors express criticism regarding the use of the word ‘partnership’. D. W. Brinkerhoff and Brinkerhoff (2011, p. 12) claim that the term ‘partnership’ has been overused which has led some researchers to dismiss it as “conceptually empty and merely politically expedient”. Similarly, Pattberg and Widerberg (2016, p. 43) state that the term ‘partnership’ has been used by researchers and practitioners to “describe just about any type of collaboration between state and non-state actors” and criticize that its reputation and informative value have suffered from conceptual vagueness and value-bound agendas of partnership promotion. Biekart and Fowler (2018, pp. 1694–1695) emphasize their choice of the term ‘initiative’ as they strive to prevent a – sometimes intentionally – misleading conception of an equal power distribution amongst the involved stakeholders, which they argue would be implied by the term ‘partnership’.

### 2.1.6 Sub-Categorizations among Multi-Stakeholder Initiatives

Apart from the different overall denominations and definitions, multi-stakeholder initiatives can be differentiated and subdivided according to main characteristics. This chapter will give an overview of some distinctions present in academic literature.

Bakker et al. (2019) distinguish between ‘certification MSIs’ and ‘principle-based MSIs’. Certification MSIs focus on compliance with a predefined set of rules and standards and a verification process of the members’ compliance, often in form of attaining a certification label upon compliance. Some examples of certification MSIs are the Fairtrade certification or the Forest Stewardship Council. Principle-based MSIs such as the UN Global Compact operate based on broader principles rather than on quite narrow rules, such as shared basic values, guiding frameworks and exchange platforms; their members do not have to commit to regulations or undergo verification.

Rasche et al. (2013) differentiate between four different types, comprising the two mentioned by Bakker et al. (2019) as well as ‘reporting initiatives’ and ‘process-based initiatives’. Reporting initiatives provide frameworks with detailed instructions and indicators to enhance the disclosure of social and environmental performance; examples are the Global Reporting Initiative or the Greenhouse Gas Protocol. Regarding the dichotomy of Bakker et al. (2019), reporting initiatives would be part of the principle-based initiatives. Process-based initiatives focus on processes around CSR and support organizations to improve their sustainability management systems. Rasche et al. (2013) indicate ISO 26000 standard as an example for a process-based initiative which indicates management procedures for the incorporation of social responsibility.

Bakker et al. (2019) on the other hand, actively exclude ISO standards from the scope of MSIs, based on the argument that those standards are being determined by technical working circles instead of a multi-stakeholder approach.

## 2.2 Scientific Literature on Multi-Stakeholder Initiatives

In the previous chapter, I have depicted the variety of terms and definitions for collective action efforts in academic literature. In this chapter, I will give an overview of the state of scientific literature on MSIs.

Along with the augmentation of multi-stakeholder initiatives in number and influence, which I have already briefly touched in the previous chapter and will further describe in chapter 2.3, scholarly interest in MSIs has grown. Bakker et al. (2019), who performed a cross-disciplinary literature review of 293 articles on MSIs for sustainability, illustrate an incremental growth of publications on MSIs since 2005. Particularly high numbers of published articles can be observed in 2012, the year of the United Nations Conference on Sustainable Development (cf. chapter 2.3.1) and around the preparation and adoption of the UN Agenda 2030 and the Sustainable Development Goals in 2015 (cf. chapter 2.3.2).

Although a vast amount of literature about MSIs has emerged over recent years, it is fragmented and scattered across various academic disciplines, making it hard to comprehensively manage, compare, and ascertain how a specific academic field can contribute to further research (Bakker et al., 2019; Biermann et al., 2007). Biermann et al. (2007) state that literature on multi-stakeholder initiatives is characterized by segmentation as well as inconsistency of conceptualizations and levels and subjects of studies. Research efforts on MSIs are dominated by single case studies and are directed towards different partnership functions (cf. chapter 2.4), geographical levels, and policy situations, which makes them largely descriptive and hardly comparable (Bakker et al., 2019; Biermann et al., 2007, p. 240; Dingwerth, 2007, p. 6). Also, those studies often do not discuss their empirical findings in relation to theoretical concepts and therefore do not contribute to theorization in terms of theory development or testing (Bakker et al., 2019, pp. 344–345).

While various scholars highlight the necessity of more research on the topic in order to improve MSIs (e.g. Bakker et al., 2019; Jastram & Klingenberg, 2018; Mena & Palazzo, 2012), research

indicates that MSI practitioners consider and incorporate scientific evidence only to a limited extent as a basis for decision-making, alongside with other sources of information, like region-specific or indigenous knowledges and grey literature (Roberts, 2019).

## 2.3 Emergence and Development of Multi-Stakeholder Initiatives

Multi-stakeholder initiatives have emerged in the 1990s and have strongly grown in number and influence since. They are broadly regarded as a solution to global governance gaps (cf. chapter 2.4) and fundamental to achieving the SDGs, but during the past years MSIs and their effectiveness have also been increasingly challenged and criticized (e.g. Bäckstrand, 2006; Biermann et al., 2007; Eweje et al., 2021; Oguntuase, 2020; Pattberg & Widerberg, 2016).
This chapter will provide an outline of the historical development and promotion of multi-stakeholder partnerships since the early 1990s as well as common points of criticism.

### 2.3.1 Origin and Rise of Multi-Stakeholder Initiatives

Multi-stakeholder initiatives started to emerge in the 1990s in the context of the 1992 Earth Summit (United Nations Conference on Environment and Development) in Rio de Janeiro (Pattberg & Widerberg, 2016, p. 43). Agenda 21, the plan of action developed and consequently adopted by more than 178 governments at the 1992 Earth Summit, appealed for and claimed to mark the beginning of a "global partnership for sustainable development" (United Nations, 1992, p. 3). Agenda 21 emphasized the importance of other organizations' contributions, in addition to the national governments and the UN system, on different geographical levels as well as public participation to sustainable development efforts. In the context of chapter 7: 'Promoting sustainable human settlement development', Agenda 21 explicitly referred to "partnerships among the public, private and community sectors" as a basis for improvement (United Nations, 1992, pp. 45–46) and indicates the creation of a policy environment enhancing such partnerships as a reinforcing activity to be taken by all countries (United Nations, 1992, pp. 61–62).

A decade later, the 'Type II partnerships', that have also become known as the 'Johannesburg partnerships', were announced as one of the main outcomes and innovations of the 2002 World Summit on Sustainable Development in Johannesburg (Hens & Nath, 2003). While public-private partnerships had become quite common on a national level since the early 1980s, they were new arrangements in transnational governance at that point (Biermann et al., 2007).

These Type II partnerships aimed to supplement – and partly substitute – the traditional inter-governmental agreements (Biermann et al., 2007; Pattberg & Widerberg, 2016) and to integrate other actors from civil society and business as actors in the implementation of the Agenda 21 objectives (Bäckstrand, 2006; Hens & Nath, 2003), hence bridging governance gaps (Biermann et al., 2007), linking multilateral agreements with local action (Bäckstrand, 2006) and facilitating implementation on all levels (United Nations, 2002). They have been presented and broadly received as a powerful and more democratic and participatory tool for the implementation of Agenda 21 on all different levels (Biermann et al., 2007; Hens & Nath, 2003).
Those Type II partnerships are often referred to as multi-stakeholder partnerships, which subsume different specifications of partnerships that are mentioned in the outcome documents of the 2002 World Summit on Sustainable Development (United Nations, 2002), such as:

- public-private partnerships
- community-based partnerships
- "partnerships among interested governments and stakeholders, including the private sector, indigenous and local communities and non-governmental organizations" (United Nations, 2002, p. 30)
- multi-sector partnerships
- "partnerships between scientific, public and private institutions" (United Nations, 2002, p. 55)
- "partnerships involving governments, international organizations and relevant stakeholders" (United Nations, 2002, p. 65)
- "partnerships between governmental and non-governmental actors, including all major groups, as well as volunteer groups" (United Nations, 2002, p. 71)

The important role of multi-stakeholder partnerships has been further emphasized during the United Nations Conference on Sustainable Development in 2012 (Rio+20). The outcome document of the Rio+20 Conference, "The future we want" (United Nations, 2012) reaffirms the role of public-private partnerships and the importance of the private sector as well as broad public participation and involvement of all stakeholders for the achievement of sustainable development, and re-commits to the global partnership for sustainable development launched at the 1992 Earth Summit.

### 2.3.2 MSIs and the Sustainable Development Goals

The establishment of the Sustainable Development Goals by the United Nations in 2015 further enhanced the development and status of multi-stakeholder initiatives. In Sustainable Development Goal 17: "Strengthen the means of implementation and revitalize the Global Partnership for Sustainable Development", the 2030 Agenda for Sustainable Development of the United Nations Agenda explicitly acknowledges the contribution of multi-stakeholder partnerships and commits to their enhancement (United Nations, 2015b).

The launch of the "Partnerships for SDGs platform" in 2015 further strengthened the position of MSIs and the efforts to reach the SDGs through collective action efforts (Global Development Incubator, 2015). As of September 2021, 5,483 voluntary commitments and multi-stakeholder partnerships had registered on the platform (United Nations). In Figure 1, the registered initiatives are sorted by the Sustainable Development Goal(s) to which they are contributing. The initiatives spread across all SDGs, with a strong spike in SDG 14 ("Life Below Water") which can be attributed to the United Nations' Ocean Conference in 2017 that led to over 1,400 voluntary commitments as a result (United Nations, 2018). Some other strongly represented Sustainable Development Goals are SDG 4 ("Quality Education"), SDG 8 ("Decent Work and Economic Growth") and SDG 17 itself, whereas SDG 10 ("Reduced Inequalities"), SDG 9 ("Industry, Innovation and Infrastructure") and SDG 16 ("Peace, Justice and Strong Institutions") are the least represented ones.

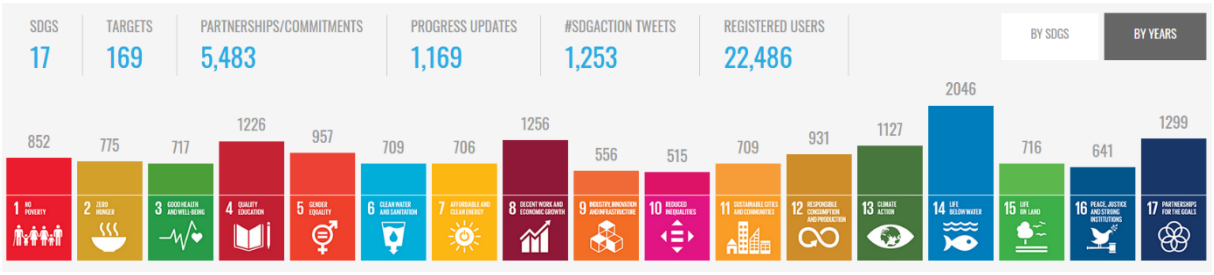


*Figure 1.* Initiatives registered on the Partnerships for SDGs online platform by SDGs. (United Nations, retrieved from https://sustainabledevelopment.un.org/partnerships/, 15.09.2021)

In Figure 2, the registered initiatives are displayed by year (the source, however, does not indicate whether the year refers to the date of foundation, registration, or else). A multiple increase in initiatives can be observed in 2012, the year of the Rio+20 United Nations Conference on Sustainable Development, and another strong rise in 2017, the year of the UN Ocean Conference. Overall, the number of initiatives has been continuously growing (with a stagnation between 2010 and 2011) over the indicated period.

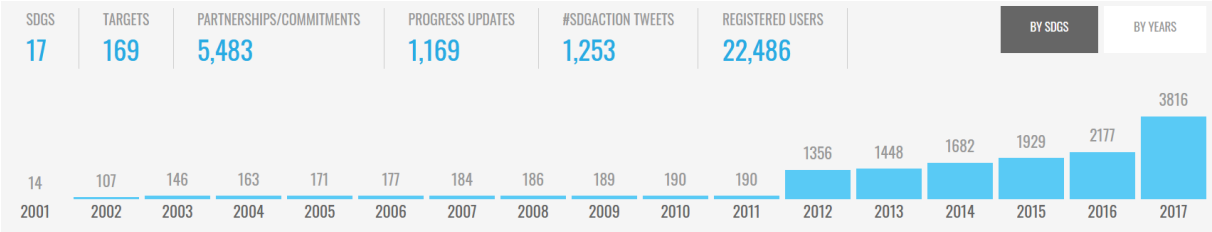


*Figure 2*. Initiatives registered on the Partnerships for SDGs online platform by year. (United Nations, retrieved from https://sustainabledevelopment.un.org/partnerships/, 15.09.2021)

The indicated number of initiatives, however, does not provide information on their state of activity. In their "Partnership Exchange 2018 Report", the United Nations (2018, p. 30) indicate that of the 3,835 initiatives registered by August 2018, 1,830 had reported the completion of their activities, amounting to approximately 48% of all registered initiatives.

Regarding the involvement of the private sector in multi-stakeholder initiatives, MSI Integrity (2020b) indicate an involvement in MSIs of over 10,000 companies, among them 13 of the 20 biggest global companies by revenue. They sustain that MSIs have been legitimized as good practice by the support of mighty governments and international corporations and the patronage of several prominent CSOs.

### 2.3.3 MSIs in Global Climate Governance

Following on what I previously described on collective action (chapter 2.1.1) and global governance (chapter 2.1.3), in global climate governance the atmosphere is a non-excludable common good, which has been traditionally used by a multitude of actors to dispose of their emissions. At the same time, the atmosphere's capacity to absorb emissions without severe modifications of the global climate is a limited resource, and there is a broad scientific consensus that we are exploiting this capacity and closing in on this limit at an increasing velocity (Jagers & Stripple, 2003, pp. 385–386).

Climate governance has been particularly readily and strongly globalized, due to several reasons which include political and institutional factors as well as the complex nature and global dynamics of climate change. Climate governance encompasses an especially wide range of affected actors and other stakeholder groups with political and economic interests on a cross-border level, which entails a high involvement of MSIs and non-state actors in general. The complexity of climate dynamics requires a coordination of policies across hierarchies, sectors, and organizational forms. The Kyoto Protocol implicated new incentives and obligations for its signatory countries to develop governance structures and market mechanisms. Regarding

important nation-state players that did not commit to the Kyoto Protocol, interests and efforts have increased to enhance their contribution to climate change mitigation in some other form such as voluntary engagement in MSIs (Andonova et al., 2009, pp. 57–58).
In climate governance, interests and efforts are rather organized around one climate-related subject such as renewable energy or emission trading than around a certain industry or group of actors, as is often the case in other areas of environmental governance (Andonova et al., 2009, pp. 57–58). Governance activities range from information and networking support, to enhancing voluntary commitments and self-regulation as well as operational activities such as running carbon emission registries or financing mechanisms (Hale & Roger, 2014, p. 60).
In congruence with the increasing urgency and the particularly global character of climate dynamics and climate change, the number of transnational initiatives for climate change mitigation has strongly increased during the past decades (Gregorio et al., 2020, p. 1; Hale & Roger, 2014, p. 72) and the position of non- and sub-state actors in climate action has been further endorsed by their recognition and promotion by the 2015 Paris Agreement (Gregorio et al., 2020, p. 1; Roger et al., 2017, p. 3). The Paris Agreement strongly builds on the strategy of orchestration, which means that international organizations and states involve and coordinate transnational partnerships, non-state and substate actors in their specific sub-ordinate targets contributing to the overall collective action goal, and has solidified the United Nations Framework Convention on Climate Change (UNFCCC) as the key orchestrator of MSI efforts (Bäckstrand & Kuyper, 2017, p. 765).

### 2.3.4 Increasing Criticism of MSIs

In this chapter, I will briefly outline the growing criticism of MSIs as part of the course of their development. In the respective subsequent chapters, I will further discuss the main points of criticism regarding the lack of legitimacy (cf. chapter 2.5.6), effectiveness (cf. chapter 2.6.3), and evaluation (cf. chapter 2.7.3).

While many scholars and practitioners treat MSIs as significant contributors to global governance and a ‘new solution’ to environmental and social problems, criticism has been increasing during the past years concerning their effectiveness as well as their directionality (e.g. Biermann et al., 2007; Pattberg & Widerberg, 2016).
MSI Integrity (2020b) describe an incremental withdrawal of civil society organizations from multi-stakeholder initiatives due to growing skepticism, resulting in increased influence and

power of the private sector in global governance affairs. Critics challenge the instrumentalization of MSIs for the expansion and exertion of already existing power by taking advantage of the United Nations' and other international organizations' auspice and 'good name' (Bäckstrand, 2006; Biermann et al., 2007; Pattberg & Widerberg, 2016).

MSI Integrity (2020b) state that

> [t]he term "MSIs," which did not have a negative connotation when it was used in the UNGPs [UN Guiding Principles on Business and Human Rights], has become increasingly connotative of a corporate-oriented model or a model that is not focused on accountability. (MSI Integrity, 2020b, p. 10)

This aligns with another point of criticism, namely the lack of accountability and legitimacy, addressed by several authors (Bäckstrand, 2006; Biermann et al., 2007).

Overall, MSI Integrity (2020b) come to the conclusion that multi-stakeholder initiatives have surpassed the summit of their significance. The authors further emphasize that MSIs are not apt and able to fill governance gaps and act as substitutes for public regulation and suggest that new worker-driven models might be more suitable for the purpose and might possibly replace MSIs in the mid- to long-term future.

## 2.4 Main Functions of Multi-Stakeholder Initiatives

Multi-stakeholder initiatives have emerged and are expected to meet deficits in global governance, which have not been covered by traditional nation-state governance and international agreements. There are three main deficit areas to cover: (1) a *regulatory deficit*, (2) an *implementation deficit* and (3) a *participation deficit*. Or, in other words, to compensate those deficits can be regarded as the three main functions of MSIs (e.g. Bäckstrand, 2012; Biermann et al., 2007).

(1) *Regulatory deficit* concerns the lack of intergovernmental regulations and agreements for certain issues of sustainable development, which MSIs are meant to fill through self-regulations.

(2) *Implementation deficit* concerns the lack or ineffective implementation of existing regulations and agreements, which MSIs should contribute to put into effect.

(3) *Participation deficit* concerns the insufficient participation of marginalized groups and key stakeholders, which MSIs are supposed to help overcome by their incorporation into decision-making processes and a balanced power distribution among different partners (Biermann et al., 2007).

## 2.5 Legitimacy of Multi-Stakeholder Initiatives

One of the two main expectations and categories of assessment of MSIs is their legitimacy (Widerberg & Pattberg, 2015, p. 47); at the same time, insufficient legitimacy of MSIs constitutes a frequent point of criticism (e.g. Bäckstrand, 2012; Biermann et al., 2007; Mena & Palazzo, 2012).

In chapter 2.5.1, I will introductorily describe why the issue of legitimacy is relevant to MSIs, which will be clarified in more depth in the explanation of the concept of legitimacy in chapter 2.5.2. Subsequently, I will characterize different dimensions of legitimacy in chapter 2.5.3 (sociological and normative legitimacy) and 2.5.4 (input and output legitimacy). Chapters 2.5.5 and 2.5.6 particularly focus on democratic legitimacy and its possible sources, and the last chapter links legitimacy to the effectiveness of MSIs.

### 2.5.1 Relevance of Legitimacy for MSIs

As mentioned in chapter 2.4, multi-stakeholder initiatives are expected to meet the regulatory and implementation voids left by national governments and intergovernmental agreements. In order to achieve this goal, actors are required to follow the established regulations and implement measures along their guidelines. As MSIs rely on voluntary compliance and largely dispose of no sanction mechanisms (except for certification MSIs which can exclude members but still are depending on voluntary compliance; cf. chapter 2.1.6), legitimacy is considered a fundamental prerequisite to fulfill their function (e.g. Bäckstrand, 2012; Gregorio et al., 2020; Mena & Palazzo, 2012; Scharpf, 2006, 2009). Or, as Bäckstrand (2006, p. 294) puts it: "Normatively, global governance can be conceived as the process of creating a legitimate political order and rule compliance in the absence of supranational authority or world government".

Jastram and Klingenberg (2018, p. 776) add that there is a broadly assumed correlation of the legitimacy and the effectiveness of an MSI, which accounts for the increasing interest in legitimacy issues of MSIs in scientific and management literature.

### 2.5.2 The Concept of Legitimacy

In polities, different mechanisms exist to incite actors to comply with the rules and policies of the system and to maintain order and social control. Three different drivers can be distinguished: (1) Coercion, (2) self-interest/inducement and (3) legitimacy (Bäckstrand, 2006, p. 294; Bernstein, 2011, p. 20). Coercion is based on the fear of consequences at non-

compliance which makes the actors obey the rules. In the second case, the actors comply with the rules because they serve their self-interest. In the case of legitimacy, the actors regard the rules as legitimate and think they should be followed (Bäckstrand, 2006). Legitimacy in global governance is crucial as the other two options are mostly not available, not desirable and/or too costly (Bernstein, 2011, p. 20) since MSIs cannot resort to long-established and legitimated democratic schemes of which nation states dispose (Mena & Palazzo, 2012).

Legitimacy refers to institutions, organizations and regulations – not to the actors – and indicates the quality of the social or political order (Risse, 2004, p. 7). Scharpf (2006, 2009) sustains that shared legitimacy beliefs are a prerequisite for effective governance. According to the author, "Socially shared legitimacy beliefs are able to create a sense of normative obligation that helps to ensure the voluntary compliance with undesired rules or decisions of governing authority" (Scharpf, 2009, p. 173). Thereby the need for legitimation increases with the extent to which the governed's own interests or preferences will be restricted or violated (Scharpf, 2006). In accordance with the three possible drivers mentioned by Bäckstrand (2006) and Bernstein (2011), Scharpf (2006) states that as soon as rules oppose self-interests of the actors, shared legitimacy beliefs are substantial for effective governance; or otherwise more extensive control and sanction mechanisms to enforce compliance with the governing body's regulations and decisions would be required.

The core legitimating arguments are all based on the premise that "legitimate government must serve the 'common good' of the respective constituency" (Scharpf, 2006, p. 2), which closely relates to the concept of collective action (cf. chapter 2.1.1).

Within the concept of legitimacy, different dimensions can be distinguished, which I will describe in the following chapters, and which are illustrated in Figure 3.

### 2.5.3 Sociological and Normative Legitimacy

There is a sociological and a normative dimension adherent to the concept of legitimacy, with the first closely related to a Weberian and the second to a Habermasean position (Bernstein, 2011; Buchanan & Keohane, 2006; Dingwerth, 2007; Keohane, 2011). Legitimacy thereby always refers to institutions or rules, not to the actors (Risse, 2004, p. 7).

An institution or organization is legitimate in the sociological conceptualization when it is socially accepted as authority and "is widely *believed* to have the right to rule" (Buchanan & Keohane, 2006, p. 405, emphasis in original). The main focus of a normative conceptualization of legitimacy, in contrast, lies on the "*acceptability* of authority" (Dingwerth, 2007, p. 14,

emphasis in original), meaning whether the institution or organization actually has the right to rule, or, in other words, is normatively justified to do so (Buchanan & Keohane, 2006; Dingwerth, 2007). Williams (2009, p. 43) describes this as "the justification of actions to those whom they affect according to reasons they can accept".

Both conceptualizations are intertwined and influence each other (Dingwerth, 2007; Scharpf, 1997). Dingwerth (2007, pp. 14–15) as well as Bernstein (2011, p. 20) include both dimensions in their delineation of legitimacy, the social acceptance as well as the normative justification. In their studies, however, they focus on the normative aspects of legitimacy, as do most scholarly discussions on the legitimacy of multi-stakeholder partnerships (Bäckstrand, 2012; Bernstein, 2011; Dingwerth, 2007).

For a more comprehensible overview of the relation between the sociological and the normative dimension to each other and in context with other distinctions within the concept of legitimacy, cf. the outline of the interrelations by Dingwerth (2007, p. 14) in Figure 3.

### 2.5.4 Input and Output Legitimacy

Another categorization between input and output legitimacy is made by various authors (e.g. Bäckstrand, 2006; Mena & Palazzo, 2012; Scharpf, 1997, 2006).

Scharpf (1997, 2006) in their studies on the legitimacy of EU governance relates this distinction closely to their conceptualization of (electoral) democracy. The author describes democracy as the aim to collective self-determination with a relation to both, the input and output dimension of a political system. The input factor of self-determination concerns the authentic representation of the governed and the possibility to hold the government accountable for it. In its tradition, it derives from the ideal of Ancient Greece's participatory democracy and would correspond to either a direct democracy or a representative government with a maximization of direct participation or representation, depending on the size of constituency. The government responds to the 'common will' of the governed as far as possible and can be held accountable for doing so. The input-oriented approach poses various potential or practical problems which have been discussed by scholars and have been partly contested with practices and key values such as discourse and deliberation, which I further describe in chapter 2.5.5 and 2.5.6.

The output factor refers to the effectiveness of the government of achieving the governed's preferences. In this tradition, the danger to the common interest by the abuse of power, either by the utilization for special interests or by corrupt governors acting in their own interest, was in focus. The governing institutions are responsible for protecting the common interest against

these threats – common prevention measures being distribution of power to separated political bodies with different election mechanisms, independent (control) institutions, veto positions, a constitution protecting basic rights etc. – and therefore their legitimacy. Furthermore, the institutions are expected to ensure the effective pursuit of the common purpose, which encompasses contradictory requirements for those two functions.
Scharpf (2006) summarizes that institutions may vary in the weight they put on input- and output-oriented mechanisms, but that all of them are regarded fundamental for democratic legitimacy and a good performance in one area cannot fully compensate deficits in another.

Bäckstrand (2006, 2012) and Mena and Palazzo (2012) draw upon that twofold notion of input and output legitimacy and examine it in the context of multi-stakeholder initiatives.
The authors describe input legitimacy in terms of procedural requirements for the decision-making processes in global governance based on core democratic values participation and inclusion, consensual orientation, transparency, and accountability (see also chapter 2.5.5), and only slightly diverge in their denomination and emphasis.
In the context of multi-stakeholder partnerships, Bäckstrand (2006, 2012) equates output legitimacy with the effectiveness of the partnership agreements, which also reflects the partnership's ability to resolve problems. The effectiveness of institutions and organizations can be divided into an 'institutional effectiveness' and an 'outcome effectiveness'. The first addresses the adequacy of the institutional frame, structures, and design for achieving the aspired outcome. The second refers to the extent to which the proposed outcomes, is to say the proposed contribution to sustainable development, are reached. Mena and Palazzo (2012) regard efficacy as only one of three factors of output legitimacy and add coverage and enforcement. Coverage refers to the quantity of actors involved in implementation and enforcement to the control and sanction mechanisms of the MSI and its environment.
With regard to the 'overall legitimacy' and the compensation of legitimacy deficits of one dimension, Bäckstrand (2006, 2012) opposes Scharpf (1997, 2006) and argues that effectiveness of partnerships (output legitimacy) can counterbalance procedural lacks (input legitimacy).

Although the authors roughly coincide in their distinctions and descriptions of input and output legitimacy, they disagree about whether the whole scope of input and output legitimacy corresponds to the democratic legitimacy of an organization or institution. While Scharpf (1997) holds that input and output legitimacy are both essential parts of democratic legitimacy, Bäckstrand (2012) argues that the input or procedural dimension corresponds to democratic

legitimacy and, furthermore, that an institution can gain additional legitimacy through achieving a strong output legitimacy, as I will describe in more detail in the following chapter.

In line with Bäckstrand (2006, 2012), Dingwerth (2007) also attributes only the input dimension to democratic legitimacy and depicts the output dimension as an additional kind of legitimacy in his outline in Figure 3. In the illustration, the author also includes 'throughput legitimacy' as another dimension, based on a distinction made by Zürn (1998). According to Zürn (1998), throughput legitimacy covers the procedural fairness of the institution's operation. Zürn (1998) therefore makes another sub-division compared to the above-mentioned authors, as e.g. Mena and Palazzo (2012) explicitly include procedural fairness in their scope of input legitimacy.

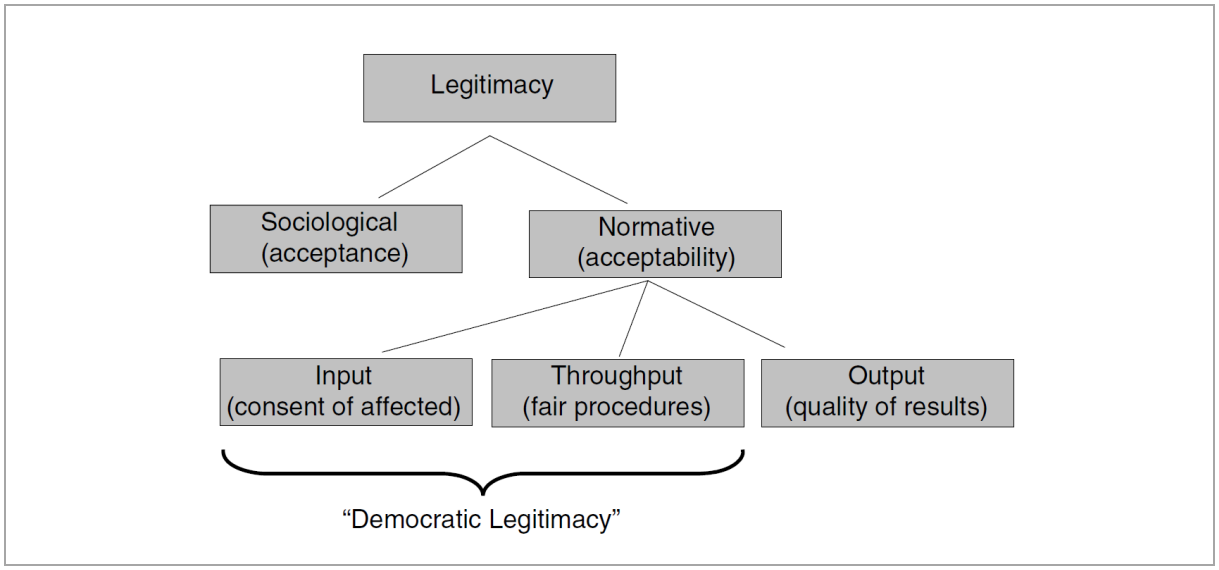


*Figure 3*. Concept of legitimacy, according to Dingwerth (2007, p. 14).

### 2.5.5 Democratic Legitimacy of MSIs

The concept of democracy and democratic legitimacy is broadly based on and tailored to nation-state government systems. Governance and regulation by democratic national governments are based on a democratic system where political decision-makers have been elected for that purpose and fulfill the claims of a direct or representative democracy and accountability mentioned in the context of input legitimacy in the previous chapter, through 'electoral accountability'. Non-state governance initiatives lack this established democratic frame but likewise face legitimacy claims. This leads to tensions and legitimacy needs for MSIs and to the question whether

democratic governance beyond national government systems is possible (Mena & Palazzo, 2012; Risse, 2004).
While Scharpf (2006, p. 4) sustains that electoral accountability is a fundamental prerequisite for all democratic polities, other authors call for innovation and adaptation of democratic legitimacy to globalization and new governance constellations (Bäckstrand, 2012; Risse, 2004).

Bäckstrand (2012) investigates the legitimacy and democratic claim of transnational public-private partnerships, focusing on the Johannesburg partnerships (cf. chapter 2.3.1). In their study, the author discusses the extent of democratic legitimacy of the Johannesburg partnerships and the applicability of the current concept of democracy to public-private partnerships.
Bäckstrand (2012) claims that public-private partnerships fall short of democratic or near democratic governance following either of the various theories, such as democratic theory, critical economy perspective or realism, due to e.g. the lack of electoral mechanisms or the sovereignty-based power structures of the global partnerships. In the expanding literature on democracy of other than nation-states, authors argue that the current concepts and mechanisms of democracy and democratic legitimacy are too strongly built on nation-states and not suitable for assessing multi-stakeholder initiatives, and should be rethought to take account of the complexity of global governance (Bäckstrand, 2012; Dingwerth, 2007; Grant & Keohane, 2005; Keohane, 2011). Keohane (2011) additionally argues that apart from adapting and developing the standards of democratic legitimacy, the thresholds for legitimacy applied in global governance structures should be lower than those applied to well-ordered national governance since no transnational institution would be able to fulfill high thresholds of liberal democracy.
As an alternative, scholars base their deliberations and discussions on the democratic legitimacy of non-state governing structures on core democratic values. The academic discussion of various authors results in similar and overlapping core principles and values, encompassing accountability, transparency, participation, inclusion, deliberation and responsiveness (Bäckstrand, 2012; Bernstein, 2011).

In the next chapter, I will discuss those principles on the basis of the extensive study on transnational rule-making by Dingwerth (2007), which encompasses most core criteria formulated by other authors.

### 2.5.6 Sources of Democratic Legitimacy

Dingwerth (2007) distinguishes three sources of democratic legitimacy: (1) participation and inclusiveness, (2) democratic control and (3) discursive practice.

*Participation and inclusiveness*: Participation constitutes a fundamental component of any theoretical approach to democracy and any decision-making process implies at least a minimum of participation. Dingwerth (2007) maintains that democratic legitimacy is therefore a question of the scope and quality of participation rather than whether there is any participation or not. Scope and quality of participation indicate who is involved in the process and how they are participating in it. As MSIs do not dispose of electoral mechanisms and representation, inclusion in this context means the equal involvement of all affected stakeholders in the decision-making process (Mena & Palazzo, 2012, p. 538), with 'affected' comprising at least everyone whose options for action will be significantly influenced by the result of the process (I. M. Young, 2000, p. 23).

Bäckstrand (2012) in this context urges to scrutinize whether participation is real or symbolic, whether stakeholders have equal opportunities to participate and in which phases of the process the participation occurs. Particularly, Bäckstrand (2012) points out three main aspects to assess participation, encompassing (1) the geographical dimension, reflected in the participation of main actors from the Global North and the Global South, (2) the dimension of non-governmental participation and (3) the dimension of participation of marginalized groups, such as women or indigenous peoples. This aspect aligns with the purpose of multi-stakeholder partnerships formulated by the United Nations and other authors of including different stakeholder groups and addressing the participation gap of traditional nation state governance (e.g. Biermann et al., 2007; United Nations, 2002; cf. chapters 2.3 and 2.4).

Concerning the geographical distribution, the Johannesburg partnerships show a geographical imbalance and predomination of actors from the Global North. Of the partnerships registered with the UN database[1], 60% include at least one partner from OECD-countries while 44% have a partner from the developing world. When we look at who leads and manages those partnerships, the imbalance becomes even more blatant. Governments leading partnerships are predominantly from industrialized countries. Among the ten most frequently leading governments

[1] During the period of writing this master thesis, the UN database of partnerships for sustainable development has been unavailable due to maintenance work, so I could not provide more recent numbers but only resort to other publications.

are only two from the developing world: South Africa and Indonesia, the countries which hosted the 2002 World Summit on Sustainable Development (cf. chapter 2.3.1) and its last preparatory conference, respectively. Regarding the country of implementation, a similar disparity becomes evident. While 70% of partnerships led by actors from industrialized nations have as country of implementation at least one OECD-country, 83% of partnerships led by actors from developing nations implement exclusively in the developing world (Biermann et al., 2007, pp. 14–15). Considering where the issues of the SDGs are most pressing further aggravates this imbalance of implementation directionality. Andonova and Levy (2003, p. 28) conclude that the Johannesburg partnerships are supply-driven instead of demand-driven and therefore are shaped by the capacities and interests of the leading and financing actors (predominantly from the Global North) rather than by the needs of the poor and/or marginalized actors (predominantly from the Global South). Such an imbalance also occurs when partnerships work on an issue of SD but omit local key stakeholders, as in the case of the EU Water Initiative with European officials working on the problem of water scarcity in sub-Saharan Africa but largely leaving out local experts and stakeholders (Mena & Palazzo, 2012; WaterAid & Tearfund, 2005).

Regarding the aspect of participation of non-governmental actors, the partnerships are dominated by governments (Bäckstrand & Kylsäter, 2014, pp. 336–337) and the participation of marginalized groups is as low as 1%, respectively, for women's groups, youth groups, indigenous peoples' groups, farmers' groups and workers and trade unions (Bäckstrand & Kylsäter, 2014, p. 337; United Nations, 2010). More institutionalized groups of non-governmental actors[2] are more numerously represented; of the registered partnerships, 9% have partners from local authorities, 18% comprise partners from the scientific and technological community, 30% have NGO partners and 38% have business and industry partners (United Nations, 2010). Within those groups, the majority of NGOs are large organizations from the Global North and small organizations and businesses are mostly absent (Bäckstrand & Kylsäter, 2014; Hale & Mauzerall, 2004). In comparison, governments are involved in 83% of partnerships, UN institutions in 62% and other IGOs in 61% (Bäckstrand & Kylsäter, 2014, pp. 336–337), and of the leading partners, 58% consist of governments or intergovernmental organization (including the UN) (United Nations, 2010).

[2] The classification of groups is based on the nine major civil society groups identified in Agenda 21 United Nations (2010).

These numbers contradict the claim that multi-stakeholder partnerships can close the participation gap (cf. chapter 2.4), particularly regarding marginalized groups. Partnerships rather seem to recreate existing disparities and mechanisms to include already powerful actors and exclude already marginalized groups (Andonova & Levy, 2003; Biermann et al., 2007). Also, the involvement of business actors has not worked out as aspired. Business actors lead only 3% of partnerships (Biermann et al., 2007, p. 16) and account for less than 1% of the partnerships' funding (Hale & Mauzerall, 2004, p. 231) although corporations had initially been among the vocal supporters of the partnership concept at the Johannesburg summit (Biermann et al., 2007; Hale & Mauzerall, 2004). Small organizations and businesses are hardly involved, neglecting the potential to include actors with detailed knowledge about local development concerns (Hale & Mauzerall, 2004).

*Democratic control*: Democratic control refers to the possibilities to achieve a high responsiveness from governors and to hold governors accountable for representing their constituencies' interests (Dingwerth, 2007) and aligns with the understanding of input legitimacy by Scharpf (1997, 2006) described in chapter 2.5.4. Dingwerth (2007) further divides democratic control into transparency, accountability, and responsiveness in a consecutive logic: Transparency helps to obtain higher accountability, and accountability contributes to attain maximum responsiveness which is the paramount target.

Transparency describes the access of all actors who might be affected by the results of the decision-making process to information about the process, its structure, status, timeline, subject etc. This includes the provision and availability of information on the one side and the resources to access and process it (infrastructure, skills etc.) on the other side (Bäckstrand & Kylsäter, 2014; Dingwerth, 2007). Reynaers and Grimmelikhuijsen (2015) further distinguish external and internal transparency. External transparency refers to all information made available to the public. Internal transparency is understood as a process to clarify and make available information about expectations, project parameters, and performance specifications to all MSI members. The authors regard internal transparency as a prerequisite for external transparency and for internally assessing whether the initiative is acting in the interest of the common good.

Accountability implies "that the actors being held accountable have obligations to act in ways that are consistent with accepted standards of behavior and that they will be sanctioned for failures to do so" and that those who hold them accountable have the right to do so and the capacities to judge their compliance (Grant & Keohane, 2005, pp. 29–39). Dingwerth (2007) points out that democratic control is easily confused with control and recommends to assess not

only whether there are control mechanisms in place but also who carries them out and who is able to access them.

Reynaers and Grimmelikhuijsen (2015) investigate the internal transparency among members of four public-private partnerships in the Netherlands. The authors conclude that internal transparency is present in the partnerships and not as complicated to achieve as previous literature supposes. The complexity to obtain internal transparency, however, increases during the operating life of an initiative and due to an increasing demand for detailed information on aspired performance levels, which are usually mor difficult to determine when they depend on qualitative data rather than on technical metrics.

In the realm of climate governance, Bäckstrand and Kuyper (2017) examine two orchestration efforts of the UNFCCC, the Lima-Paris Action Agenda and the Non-state Actor Zone for Climate Action, with regard to their democratic legitimacy. In both cases, affected individuals have very limited access to information about performance, effectiveness, and value-added and hence can hardly track decisions and their origin. In conclusion, the authors state that the mechanisms for transparency and accountability are "nascent at best, nonexistent at worse" (Bäckstrand & Kuyper, 2017, p. 785).

Hale and Mauzerall (2004)[3] conducted a large sample size analysis of the external transparency of 250 partnerships registered with the UN database, on the basis of three criteria: web site, reporting system and monitoring. The authors find that 64% of partnerships run a web site, 69% have some sort of reporting system in place and less than 50% run a monitoring system. The authors argue that partnerships are only appropriately transparent when they couple an effective reporting and monitoring system and provide the respective information publicly. Although nearly 90% of partnerships cover at least one criterion, only 27% meet all three criteria. These percentages can be attributed to a structural problem accounting for the "chronic lack of – and need for – stronger transparency and accountability" (Bäckstrand & Kuyper, 2017, p. 781): Reporting is largely voluntary and based on self-description (Bäckstrand & Kuyper, 2017) and the response by the partnerships is limited (United Nations, 2010, p. 23). Transparency guidelines – despite gradual improvements – are weak, mandatory requirements and the possibility of enforcement are largely absent, and therefore accountability is limited (Hale & Mauzerall,

[3] As I could not find a more recent analysis of MSI transparency with an extensive sample size, I provide the results of this publication even though it is not very new.

2004). Although the theoretical possibility exists, no Johannesburg partnership has been ever excluded from the registry due to deficient performance (Bäckstrand, 2012, p. 174).

*Discursive practice*: The quality of discourse and deliberation processes relates to a number of factors such as the rationality of discourse, inclusiveness and power relations of deliberation processes, impartiality and respect in discourse ('reciprocity') and consensus-orientation (Bäckstrand, 2012; Dingwerth, 2007; Mena & Palazzo, 2012). In MSIs, various stakeholders with different power positions and inherent power imbalances come together. Those resource and power imbalances among stakeholders might result in the disregard of less resourced partners standpoints or even their co-option by larger partners (Hale & Mauzerall, 2004). MSIs are therefore required to design their decision-making processes in a way that neutralizes those power differences and enables an equal participation in discourse and deliberation of all actors (Mena & Palazzo, 2012).

Bäckstrand (2012) criticizes that previous studies have investigated deliberation about partnerships, but not deliberation within partnerships. Deliberation processes about the Johannesburg partnerships also have a more institutionalized frame through the annual partnership fairs organized by the UN (Bäckstrand, 2012; United Nations, 2010). The quality of discourse and deliberation is closely linked to the quality of participation and is hindered by barriers to participate and by power disparities (Bäckstrand, 2012). Much of the data previously provided regarding participation and inclusiveness is therefore also very relevant for the deliberative practice in partnerships.

The three sources of democratic legitimacy established by Dingwerth (2007) cannot all be pursued independently to a maximum, as they are interrelated, partly overlapping and sometimes opposing each other. For example, tensions exist between the demands for inclusiveness and for the success of consensus-oriented deliberative processes, or between the claims for transparency and for efficiency. Therefore, those criteria do not serve as absolute standards in practice and trade-offs are required (Dingwerth, 2007; Gregorio et al., 2020).

### 2.5.7 Legitimacy through Effectiveness

As described in chapter 2.5.5, Bäckstrand (2012) sustains that democratic legitimacy of the Johannesburg partnerships is low or mixed at best, due to the deficits in participation and

inclusion, the absence of effective transparency and accountability mechanisms and the power imbalances in discursive practice described in the previous chapter. Partnerships can, however, additionally attain legitimacy through another source than democratic values, which is effectiveness. As mentioned in chapter 2.5.4 on input and output legitimacy, output legitimacy refers to the effectiveness of an MSI or institution, regarding its capacity to reach the 'best interest' of the actors and for problem-solving, and is detached from democratic legitimacy and considered another source of legitimacy by several authors (Bäckstrand, 2012; Dingwerth, 2007).

Bäckstrand (2012) claims that a high output legitimacy and therefore effectiveness can compensate for a weak input legitimacy (cf. chapter 2.5.4), arguing that if partnerships possess a strong input or procedural democratic legitimacy but fail to contribute to the collective SD problems, their legitimacy would be challenged.

Effectiveness will be subject to the next, separate chapter, as it can be regarded a source of legitimacy but above all has a fundamental importance in successfully addressing collective action problems such as climate change.

## 2.6 Effectiveness of Multi-Stakeholder Initiatives

The effectiveness of MSIs constitutes, along with legitimacy, the other main expectation and category of assessment of MSIs (Widerberg & Pattberg, 2015, p. 47). But the conceptualization of effectiveness is heterogeneous, as I will outline in the following paragraphs, where I will also argue which concept I will center on in this master thesis.

As O. R. Young (1999) writes in their introduction to the definition of effectiveness of environmental regimes,

> [a]t first glance, the meaning of effectiveness with regard to international environmental regimes seems intuitively obvious. Regimes arise to solve problems. Accordingly, effectiveness is a measure of the extent to which these arrangements succeed in solving the problems that lead to their formation. (O. R. Young, 1999, p. 109)

After some reading of academic literature on this topic, however, the issue reveals to be more complex and ambiguous, ranging from different concepts of 'effectiveness' to differing denominations of similar concepts (Underdal, 2004, p. 28).

After a period of dispersed terminology and conceptualization, several authors and publications have revisited and established David Easton's three dimensions of effectiveness (Easton, 1965) by the denomination of output, outcome, and impact effectiveness in the context of global governance and MSIs, which are based on the concept of regime effectiveness (Bäckstrand, 2012;

Barkemeyer et al., 2015; Göbel, 2010, p. 41; Miles et al., 2002; Underdal & Young, 2004; O. R. Young, 1999; Zelli et al., 2020).

Against this background, I will focus on that same concept of effectiveness in this master thesis, being aware that other conceptualizations also exist. The three mentioned dimensions of effectiveness have their origin in the concept of regime effectiveness, which again is based on regime theory. As a basis to comprehend the concept of regime effectiveness and its three dimensions, I will shortly outline the underlying regime theory in chapter 2.6.1. In chapter 2.6.2, I will explain the conceptual frame of regime effectiveness and describe its three dimensions in more detail in chapter 2.6.3. In section 2.6.4, I will briefly outline additional approaches to effectiveness.

### 2.6.1 Regime Theory

The concept of regime effectiveness is based on regime theory, which is an approach within the discipline of political sciences. Regime theory was initially used to explain cooperation between states, focusing on their relevance for avoiding international anarchy and confronting collective action problems (Bradford, 2007). The concept has since been further developed to accommodate the role and attributes of non-state actors (Bradford, 2007; Meiches & Hopkins, 2012). Analogous to collective action problems, states can be regarded as unitary utility maximizers which in interdependent situations can enter cooperations to obtain joint benefit but also have motivations to defect (O. R. Young, 1999, p. 190).

Regimes are cooperative arrangements (Miles et al., 2002, p. 5), either between states, or, on a more modern notion, between states and/or non-state actors (Bradford, 2007; Meiches & Hopkins, 2012). They can be operating in different areas of international politics, e.g. security, human rights or environment (Bradford, 2007). Regarding a specific definition of ‘regime’, Stephen Krasner’s characterization has become canonical (Steffek et al., 2021, p. 36): “Regimes can be defined as sets of implicit or explicit principles, norms, rules, and decision-making procedures around which actors’ expectations converge in a given area of international relations” (Krasner, 1982, p. 186 as cited in Steffek et al., 2021, p. 37). Consequently, regime theory holds that state and non-state actors cooperate successfully when and to which degree they achieve to establish institutional arrangements or a shared array of principles, rules, and procedures (O. R. Young, 1999, p. 191).

Regime theory seeks to explain the successes and failures of actors constituting international regimes in order to address collective action problems (O. R. Young, 1999, p. 191). Its strengths comprise the adaptability of the concept which can be and has been applied to analyze a wide range of topics (Meiches & Hopkins, 2012).

### 2.6.2 The Concept of Regime Effectiveness

Generally speaking, a "regime can be considered effective to the extent that it successfully performs a certain (set of) function(s) or solves the problem(s) that motivated its establishment" (Miles et al., 2002, p. 4). This description is not, however, a sufficient frame for an analytical assessment of effectiveness.

In order to evaluate the effectiveness of a regime, methodology necessitates a comparison of the regime against some standards of fulfillment. From these methodological requirements result three questions which need to be covered by any conceptual framework: (1) The first one concerns the exact delimitation of the object that will be evaluated. (2) The second question regards the standard against which the object will be compared. (3) The third point to be determined is how this comparison is operationalized (Miles et al., 2002).

*The object to be evaluated* can be the regime itself. In this case, only the effects of the cooperative arrangement itself will be considered and the 'gross' benefits will be determined. Alternatively, also the cost and positive side effects incurred in the process of creating and maintaining it can be considered and the 'net' benefits calculated. In that case, in addition to the regime itself, the problem-solving efforts can be evaluated (Miles et al., 2002, p. 5; Underdal, 1992, p. 229). Both costs and positive side effects can be significant. Transnational negotiations often imply big learning experiences and might, beyond multilateral agreements, lead to unilateral changes in behavior. Problem-solving efforts engender their own consequences which might even exceed the consequences of the formal agreements and regime itself. They are, however, complex and potentially costly to ascertain. The choice of the object to be evaluated is therefore of significant importance (Miles et al., 2002, pp. 5–7).

*The standard against which the object is to be evaluated* comprises a point of reference for comparison as well as a metric of measurement.

Regarding the point of reference, there are two main complementary approaches: One is to compare the effects of a regime against a fictional state where the regime does not exist. In this case, the relative improvement is determined and can provide information about whether and to what degree a regime makes a difference. The other one is to compare the effects of a regime against a satisficing situation – which might be conceptualized differently by each stakeholder – or a situation of maximizing the collective benefit (Miles et al., 2002, pp. 7–8). This approach provides information about the degree to which the addressed collective action problem actually gets resolved by the regime. It also opens up the question of what the maximum possible contribution would be to solving the collective action problem and if they regime is reaching its maximum potential. In practice of multi-stakeholder negotiation and decision-making processes, though, an optimum contribution will usually not be attained (Miles et al., 2002, pp. 7–10; Underdal, 1992, pp. 230–234). Miles et al. (2002, p. 8) visualizes the two approaches to the point of reference in relation in a graphic depicted in Figure 4.

For the metric of evaluation, the effects are usually 'translated' for example into economic criteria or biophysical criteria, such as emission values. Only the results of evaluations that used the same metric are comparable without mayor restrictions (Miles et al., 2002, p. 10).

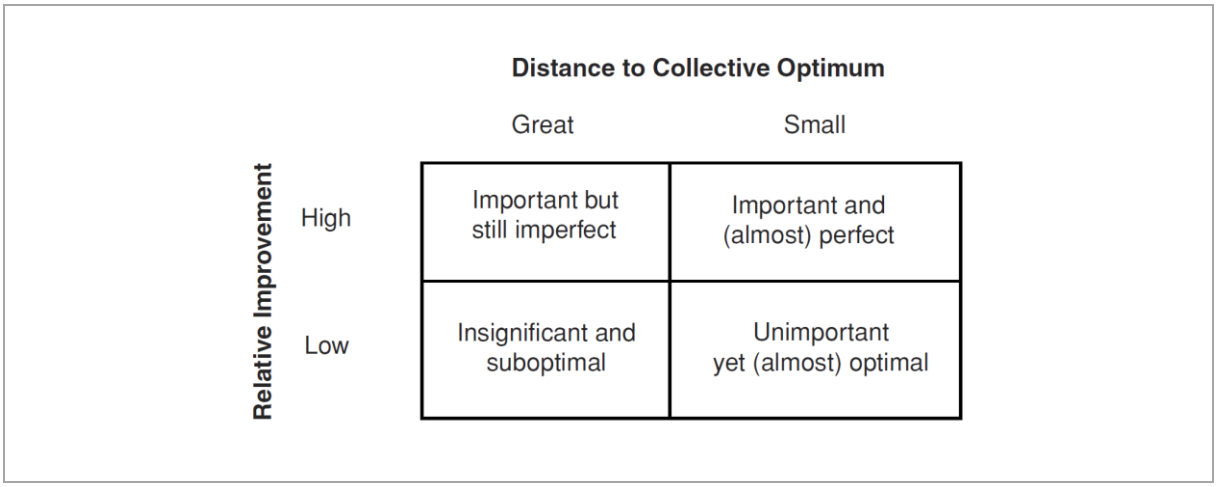


*Figure 4*. The two standard setting approaches for effectiveness evaluation in relation. (Miles et al., 2002, p. 8)

*The operationalization of attributing a score to a regime* is, according to Underdal (1992, pp. 235–237), at the same time very complex and methodologically challenging, and a problem of second order as long as no common concept of effectiveness has been established. Dissecting the methodological problems goes beyond the scope of this master thesis, so I will not go into more detail here[4].

[4] For more information on that topic, I can recommend Miles et al. (2002) and Underdal and Young (2004).

Regarding environmental regimes, the main relevance is the final improvement of the environment itself, which is attained through alterations of human behavior (Miles et al., 2002, p. 11). The relation between norms and regulations (output), behavioral changes (outcome) and biophysical ameliorations of the environment (impact) are further described in the next chapter.

### 2.6.3 Output, Outcome, and Impact Effectiveness

Originating from regime theory, three dimensions of effectiveness of governance can be distinguished: (1) *output*, (2) *outcome* and (3) *impact*, which correspond to different levels of analysis and evaluation. *Outputs* are (self-)commitments of MSIs or MSI members corresponding to aspired results. Those commitments can be held to the standards determined by certification MSIs, such as the Fairtrade certification standards, or to the principles of a principle-based MSI, such as the UN Global Compact (for certification MSIs and principle-based MSIs, see chapter 2.1.6). *Outcomes* refer to changes in human behavior that occur on the basis of said commitments. *Impacts* describe the contribution to solving collective action problems and to attaining the SDGs (Jastram & Klingenberg, 2018, pp. 776–777; Wolf, 2010, p. 4). Those three dimensions can be grouped in different ways. While outputs and outcomes involve the (self-)commitment of MSIs, the dimension of impact indicates their effects (Wolf, 2010, p. 4). Or, output is the formal result of an MSI-creating process in terms of principles and rules, while the other two are consequences of it, outcome in form of changes in behavior and impact in form of changes of the environment itself (Miles et al., 2002, pp. 5–6). Either way, the three dimensions are closely linked and interdependent and are analytically considered as a causal chain (Miles et al., 2002, pp. 6–7; Wolf, 2010, p. 4), which is depicted in Figure 5.

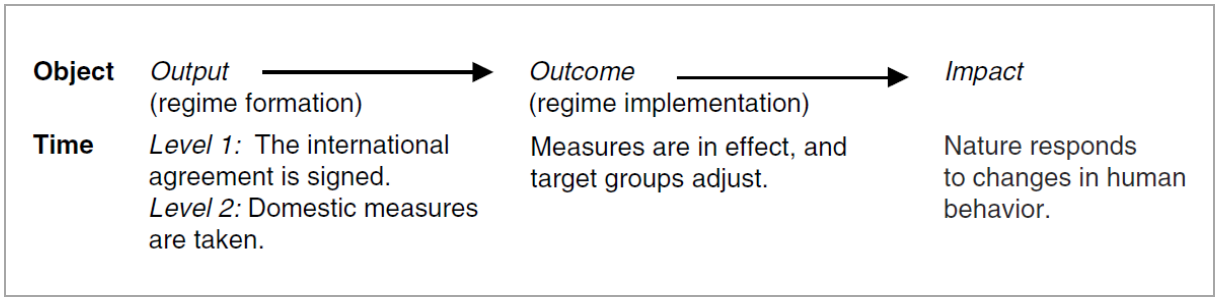


*Figure 5*. Causal chain of effectiveness dimensions for transnational environmental regimes. (Miles et al., 2002, p. 7)

*The output dimension* is the easiest one to analyze, yet gives no information on the success towards the main goal of SD. It does, however, disclose whether clear policies or regulations

exist towards goals of SD and therefore provides information about success criteria. Institutional effectiveness also lies within the realm of the output dimension, for instance the effectiveness of a monitoring system. It is in the context of the output dimension that the factors of input legitimacy of MSIs can be assessed, for example regarding the participatory or discursive quality, transparency, monitoring as well as accountability mechanisms and the responsiveness to relevant problems. Also the scope of commitment can be examined in this dimension, such as the geographical scale (local, national, transnational) and the constellation of involved stakeholders, which again relates closely to the participatory quality (Wolf, 2010). In analytical terms, output information is comparably easy to assess via document analysis with the necessary output data relatively easily accessible via webpage, sustainability reports etc., but is limited to formal commitments (Jastram & Klingenberg, 2018, p. 777; Wolf, 2010, p. 5).

Pattberg and Widerberg (2016) analyzed 340 transnational multi-stakeholder partnerships which had been active for at least five years. The authors compared those partnerships' outcomes to their publicly stated objectives. Of the investigated sample, approximately 38% were inactive and/or showed no measurable output. Another 26% showed output that did not match their declared objectives. Of the remaining 36% of multi-stakeholder partnerships, 12% showed output that partly coincided with their self-reported function, and in only 24% of the cases the output was fully congruent with the self-reported objectives, as can be seen in Figure 6 (Pattberg & Widerberg, 2016, p. 45).

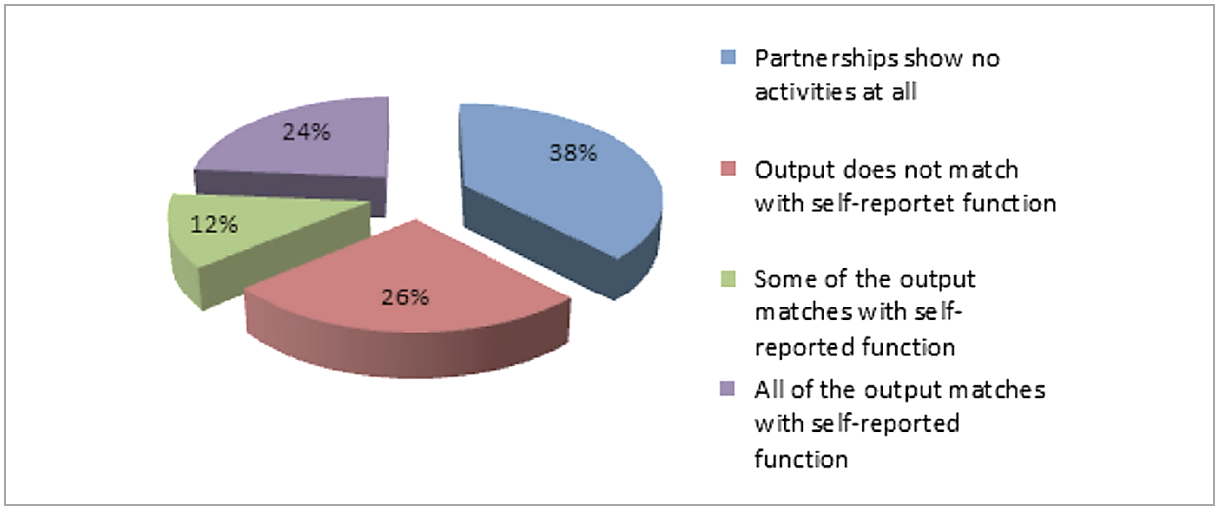


*Figure 6*. Multi-stakeholder partnership output compared to publicly stated goals and ambitions. (Pattberg & Widerberg, 2016, p. 45)

The Global Development Incubator (2015) also analyzed 17 multi-stakeholder initiatives of diverse sectors, sizes, and ages with regard to the congruence of their initial objectives and achieved results. In interviews conducted with the sponsor representatives and CEO-type representatives, the MSI initiators ranked their MSI with an average performance rating of C (out of A to F) in comparison to the initial objectives, as can be seen in Figure 7.

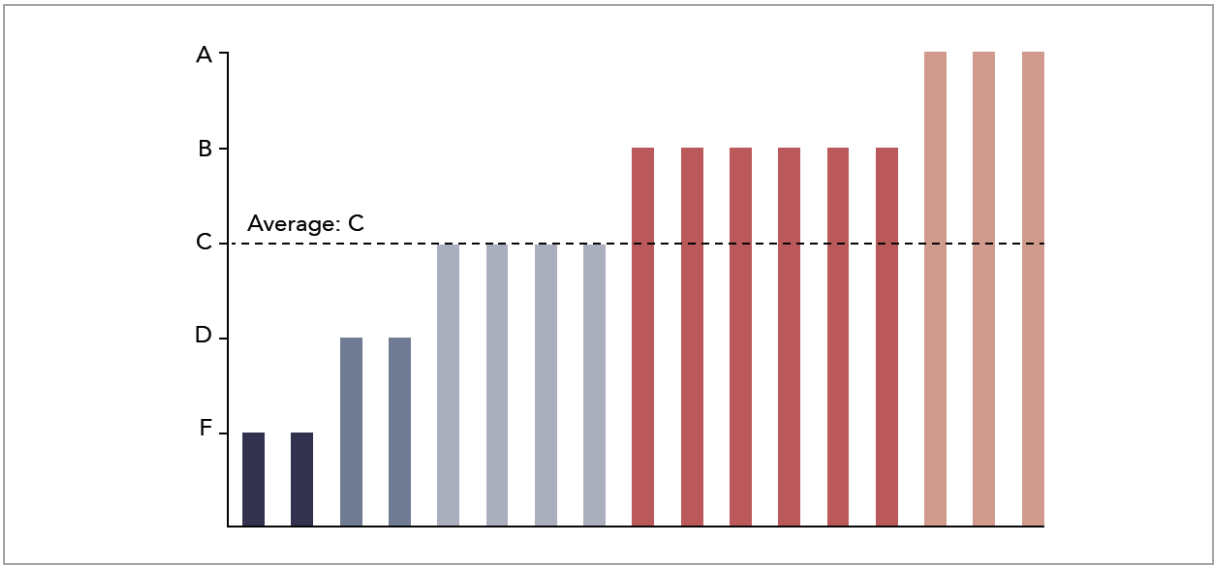


*Figure 7.* Average performance rating of MSIs compared to initial objectives (scale of A to F). (Global Development Incubator, 2015, p. 9)

Barkemeyer et al. (2015) investigate the effectiveness of the Global Reporting Initiative (GRI) which aims at promoting the disclosure and transparency of impacts on SD and at enhancing companies to take responsibility for their impacts by providing sustainability reporting standards (Global Reporting Initiative, 2022). For that purpose, Barkemeyer et al. analyzed a sample of 933 GRI reports by companies across countries, continents and industries. The GRI showed to be very successful in terms of output effectiveness, particularly regarding the engagement of Asian and South American companies (Barkemeyer et al., 2015, p. 313). This did not apply, however, to the other two effectiveness dimensions, as I will discuss in the respective section.

*The outcome dimension* goes beyond formal commitments and instead examines their concrete fulfilment and behavioral actions. “The outcome dimension builds the conditional bridge between outputs and impacts” and constitutes the “primary requirement without which impacts cannot be realized” (Jastram & Klingenberg, 2018, p. 777). It encompasses management reactions and behavior along the made commitments in order to consequently achieve a positive impact on collective action issues. Outcomes can be conceived as “measurable changes of

behavior caused by a governance initiative such as a MSI" (Jastram & Klingenberg, 2018, p. 777). Its analysis is based on data on managerial decisions and actions which is less easily accessible, which is why the outcome dimension is far less scientifically investigated in comparison to the output dimension. Another challenge is to analyze those changes in behavior, not the behavior itself and to identify whether these changes have been initiated by the MSI or do not correlate (Jastram & Klingenberg, 2018).

Jastram and Klingenberg (2018) in their study on outcome effectiveness among members of the participants of the UN Global Compact Network German found behavior changes on corporate management level in 75% of the cases (with some reservations as subsequent qualitative interviews showed that some of the changes might rather be attributed to alterations in structures or communication of already established initiatives instead of the adoption of the UN GC standards). Regarding the employee level and strategic monitoring, a deficit of measures became apparent. The authors conclude that serious top management commitment and implementation management together with strong strategic leadership and internal communication efforts are as important as external communication and lobbying with strategic decision-makers, but that those aspects are frequently underrated by MSI practitioners.
Barkemeyer et al. (2015) in their analysis of the GRI conclude that the initiative's outcome effectiveness is limited, as the content was uniform across the report samples and hardly included any geographical an sectoral differences. The reporting therefore appears to fail to reflect materiality considerations and to improve the interaction of stakeholders and companies based on the reported information.

*The impact dimension* is the dimension with the highest political relevance. An ascertained positive impact would make the strongest case for MSIs in global governance for SD. It is at the same time the most complicated dimension to analyze. The first difficulty of analysis is the question what the relevant impacts are, direct and indirect ones. Another analytical problem is who defines the relevant goals and criteria and who has the capacity and entitlement to assess them (Wolf, 2010, p. 6). Wolf (2010) suggests a possible evaluation of goal attainment based on three different groups of actors: (1) On 'corporate satisfaction', evaluating the achievement of the goals and commitments adopted by a member of an MSI along self-defined criteria. A downside to this approach is that the self-set goals might on the one hand be unambitious and on the other hand potentially do not reflect the interests of the affected stakeholders. (2) A more comprehensive approach is based on 'stakeholder satisfaction', which possibly is the most

challenging to examine and at the same time the most balanced one. In this context the question of quality of participation arises again; in order to be truly fair and balanced, all affected stakeholders would need to be included. (3) Ultimately, evaluation can be based on 'researchers' satisfaction', referring to scientific analysts' conceptualization of the collective interests.
A comprehensible evaluation of impact effectiveness needs to encompass all three perspectives. In MSIs, the first two approaches can be congruent, as the goal- and principle-setting of MSIs should take place with participation and consideration of the stakeholders (Wolf, 2010).
With regard to climate governance, Ven et al. (2017) argue that the impact dimension should not only cover quantifiable, immediate GHG emission reduction, as is often the case (e.g. Widerberg & Pattberg, 2015), but consider also indirect and non-quantifiable impacts that MSIs in climate governance have on the transformation and thus disrupting carbon lock-in within key institutions. To assess the transformative impact of an initiative, also the scaling and entrenchment of an intervention, the development of ripple effects, as well as feedback – both, positive and negative – within and across systems, need to be taken into account (Ven et al., 2017).

Ven et al. (2017) applied their approach to assess the impact of the Carbon Trust. The Carbon Trust developed a standard and label for determining the carbon footprint of products and aimed at contributing to decarbonization by making products with small carbon footprints identifiable to consumers and incite climate-friendly consumer choices, similarly to other certification systems like the Fairtrade label. Valued in terms of output and outcome, the Carbon Trust did not reach its self-declared targets of product labels nor of igniting shifts in consumer behavior towards low-carbon goods. Also regarding its quantifiable impact measured in GHG emission reductions, the Carbon Trust failed to meet the target values and reports little evidence of any emission reductions resulting from the product carbon footprinting. It did, however, have an indirect impact on decarbonization, when scaling and entrenchment are taken into consideration, through enhancing the dissemination of carbon footprinting and the development of other labeling standards and the identification of GHG emissions along the supply chain and subsequent changes in business behavior. Ven et al. (2017, p. 14) conclude that impact can occur beyond the defined targets and, regarding climate governance, might not be covered by emission metrics.
Regarding the impact effectiveness of the Global Reporting Initiative, Barkemeyer et al. (2015) in their study conclude that due to the lack of outcome effectiveness (more meaningful communication of stakeholders and companies), the GRI does not considerably contribute to problem-solving in the fields the company reports on and hence also lacks impact effectiveness; this

reflects the interrelation and causal chain of the effectiveness dimensions shown in Figure 5. Although the GRI has a high output effectiveness and influence, due to its very limited outcome and impact effectiveness, it is inaccurate and does not meaningfully contribute to solving the collective action problem at hand (Barkemeyer et al., 2015, p. 313).

In Table 2, the three effectiveness dimensions (output, outcome, impact) are related to their respective analytical accessibility and political relevance. They generally are counter-directed: The output dimension with the best analytical accessibility has the least political relevance; the impact dimension which is the politically most relevant one at the same time is little accessible for analytical examination. If the choice is between evaluating the effectiveness of one of the three dimensions, Wolf (2010), who discusses them in the context of corporate engagement in conflict zones, recommends focusing on the outcome dimension as a result of balancing political relevance and analytical accessibility: Outcome is closer to impact than output and even poses a precondition, at the same time it offers a less complicated frame for the operationalization of evaluation than impact. The author adds, however, that this is a pragmatic approach that prioritizes analytical accessibility over political relevance.

| **Effectiveness dimensions** | | **Analytical accessibility** | **Political relevance** |
|---|---|---|---|
| *Output dimension* | Words:<br>(Self)-commitment | Very high | Low |
| *Outcome dimension* | Action:<br>Change of behavior | Moderately high | Substantial |
| *Impact dimension* | Effect:<br>Goal attainment / Problem-solving | Low | High |

*Table 2*. Matrix of effectiveness dimensions according to analytical accessibility and political relevance. (Based on Wolf, 2010, p. 7; with my own modifications)

### 2.6.4 Other Classifications of MSI Effects

There are other ways to theorize and categorize results and effectiveness of MSIs. Bakker et al. (2019) for example establish impact as an overall term for the effects of MSIs and subdivide it into outputs and outcomes. Outputs in this case refer to results measured in numbers, such as number of certified facilities or volume of production. Outcomes describe improvements regarding the social or environmental situation. In a rough comparison, both understandings of outputs are similar, while the concept of outcomes used by Bakker et al. (2019) would rather

correlate with the impact dimension of Jastram and Klingenberg (2018) and Wolf (2010), and Bakker et al. (2019) do not separately discuss behavioral changes.

Clarke and MacDonald (2019) differentiate plan outcomes, process outcomes, and partner outcomes. In this distinction, plan outcomes relate to the collective action goal, such as the reduction of GHG emissions in the community and would coincide with the impact dimension as described in the previous chapter. During implementation, process outcomes are generated, ranging from collective learning to strategic budget planning. Partner outcomes describe the effects for the individual partners, such as higher reputation or increased expertise. These two outcome types are differentiated by where they occur/are experienced within the MSI, an each partly encompasses output and outcome components in relation to the dimensions described in the previous chapter.

With these examples, I want to illustrate that several conceptualizations of the effects of MSIs exist. They do, however, mainly cover the same aspects, naming them differently or categorizing them along different criteria.

In this master thesis, I will resort to the distinction of output, outcome and impact effectiveness as used e.g. by Jastram and Klingenberg (2018) and Wolf (2010) when referring to effectiveness dimensions, as it offers an appropriate level of differentiation and demarcation for the purpose of this work.

## 2.7 Evaluation of Multi-Stakeholder Initiatives

MSIs are supposed to meet complex global challenges and compensate deficits of traditional government structures. They face high expectations and demands regarding their basic functions, legitimacy, and effectiveness which I have delineated in the previous chapters. To assess whether they actually fulfill those requirements and contribute positively to global governance, MSIs need to be evaluated.

In this chapter, I will first indicate some possible definitions of evaluation in chapter 2.7.1, and then discuss the relevance of evaluation specifically for MSIs in chapter 2.7.2. In chapter 2.7.3, I will outline the status quo of the conduction and evidence of MSI evaluation. Concluding, chapter 2.7.4 will address evaluation frameworks and briefly introduce some existing tools and frameworks.

### 2.7.1 Definitions of Evaluation

There are different approaches to evaluation regarding its scope and orientation (e.g. Guerra-López, 2008; Stadtler, 2016; Stufflebeam, 2001), and no common definition of it exists (Gullickson, 2020, pp. 1–2). While definitions of evaluation in dictionaries and by some researchers focus on evaluation primarily as the judgement of value or quality of something and regard this judgment its single purpose, other scholars hold that this judgement serves the ultimate purpose of evaluation, namely, to provide a sound basis für well-informed and data-driven decisions and consequently the improvement of a program or solution (Cambridge University Press, 2014; Guerra-López, 2008; Gullickson, 2020; Merriam-Webster, Inc., 2022).
Guerra-López (2008, p. 6) refers to Egon Guba's claim that evaluation's main purpose is to improve rather than to prove, and all evaluation efforts should be directed to that purpose. The author states that

> [a]t its core, evaluation is a simple concept:
>
> - It compares results with expectations.
> - It finds drivers and barriers to expected performance.
> - It produces action plans for improving the programs and solutions being evaluated so that expected performance is achieved or maintained and organizational objectives and contributions can be realized. (Guerra-López, 2008, p. 6)

There are, however, numerous approaches to evaluation, some of which do not focus on achieved performance or objectives (Guerra-López, 2008, p. 6; Stufflebeam, 2001).

### 2.7.2 Relevance of Evaluation for MSIs

Evaluation is important for MSIs, on the one hand to assess their performance in general, on the other hand to determine the value added of MSIs. As they have been praised for filling global governance gaps (cf. chapters 2.3 and 2.4), their evaluation serves to critically examine whether they actually complement traditional international agreements and meet the governance deficits (OECD, 2008; Stadtler, 2016). Another reason to evaluate MSIs is to improve their set-up and operation in order to maximize their impact. In addition, it helps to gain information and insight on factors influencing success or failure of MSIs (OECD, 2008, p. 12). Ultimately, MSI members are accountable for the use of their resources (e.g. public institutions to the taxpayers, corporations to shareholders) and evaluation can serve as proof or assessment of it (OECD, 2008, p. 12; Stadtler, 2016, pp. 71–72). Additional motivation for MSI evaluation

include learning from evaluation results as basis for capacity building, enhancing ownership, commitment and motivation to continue activity (OECD, 2008, p. 12).

### 2.7.3 State of Evaluation of MSIs

Few large-scale evaluations of MSIs exist and most research has been conducted on single-case studies (Biermann et al., 2007, pp. 1–4; Stadtler, 2016, pp. 74–75). Most frequently, studies focus on output numbers, such as the amount of provided products, the number of provided certifications etc. (Stadtler, 2016, pp. 74–75). The outcome dimension is less often investigated (Jastram & Klingenberg, 2018, p. 777) and little empirical evidence exists of the actual impact on the addressed collective action problem (Stadtler, 2016, pp. 74–75).

The OECD (2008, p. 13) has conducted a survey amongst partnerships registered in the UN database with a strong environmental focus regarding their evaluation practices. Of the over 100 addressed partnerships, 32 answered the questionnaire. Of these 32, only 9 had completed an evaluation and only 5 included completed evaluations or specific plans for evaluation. 23 partnerships had not conducted an evaluation at the point of survey, of which 17 partnerships had plans to evaluate in the future and 6 had not planned for evaluation at all, which indicates that evaluation is not an established and integral part of MSIs so far. Bäckstrand (2006) upholds that the Johannesburg partnerships mostly escape evaluation of their implementation and performance because they were decoupled from transnational agreements on sustainable development from the beginning.

Scholars criticize that while multi-stakeholder initiatives are broadly perceived as important contributors to pressing global challenges and are increasing in popularity, (1) studies show little evidence to confirm their positive performance and (2) only few studies systematically evaluate the existent evidence of positive performance, when evaluation and evidence-based assessments would be a prerequisite for designing or improving MSIs (Biermann et al., 2007, p. 3; Pattberg & Widerberg, 2016, pp. 42–43).

Research results regarding the evaluation of single criteria and dimensions of legitimacy and effectiveness have already been delineated in the respective chapters. With respect to the overall functions of MSIs (cf. chapter 2.4), several researchers come to the conclusion that MSIs do not compensate the governance deficits (e.g. Bäckstrand & Kylsäter, 2014; Biermann et al., 2007; MSI Integrity, 2020b).

As for MSIs in general, the record and results of performance evaluation of initiatives for climate governance and climate change mitigation are mixed and it is an open question whether they are suitable and able to meet the deficits of traditional state governance and provide significant solutions for the collective action problem of climate change (Gregorio et al., 2020; Hale & Roger, 2014; United Nations, 2014).

### 2.7.4 Evaluation Tools and Frameworks

There are different tools and frameworks for the evaluation of MSIs. Several frameworks have been suggested for the evaluation of MSIs, focusing on the relationship of its partners, the outputs and/or outcomes or the order of effects (Stadtler, 2016, pp. 71–72). They cover some evaluation aspects and are referred to for guidance input for MSI evaluation but no fully comprehensive and broadly established framework yet exists (e.g. OECD, 2008, pp. 13–14). Widerberg and Pattberg (2015, p. 47) argue that the categorization of criteria into the realm of legitimacy and the realm of effectiveness has resulted in too narrow analyses that often focus either exclusively on technical or normative aspects. Requirements and challenges of MSIs exceed traditional governance structures and agreements and add additional complexity, and various scholars call for a more comprehensive and stakeholder-oriented evaluation approach in order to accommodate the complexity of MSIs and their manifold interrelations with the environment (e.g. Cabaj, 2014, p. 112; Liu et al., 2018, p. 1143; Stadtler, 2016; Widerberg & Pattberg, 2015, p. 47).

In the following, I briefly introduce some of the current evaluation tools and frameworks and delineate what they do and do not encompass, illustrating their variety in terms of scope, structure, target audience, partnership phase, etc.
Which of the currently existing tools were then selected for analysis and according to which criteria is explained in chapter 3.4.1.

**Africa Liaison Program Initiative's Partnership Assessment and Monitoring Tool**

The Partnership Assessment and Monitoring Tool developed by the Africa Liaison Program Initiative aims to evaluate the quality of MSI relations from the members' perspective in the course of a quick and simple-to-perform self-assessment. It is based on twelve core principles and offers an assessment matrix, including scoring specifications, for each of them. Upon field

trial, it received positive remarks on its learning potential and space to address issues regarding the functioning of the initiative (Africa Liaison Program Initiative, n. d.).

**Brinkerhoff's Assessment Framework**

J. M. Brinkerhoff (2002) proposes a framework for the evaluation of MSIs that encompasses not only the outcomes of partnerships but also addresses the partnership relationships. The author suggests methodologies of data collection for each point of evaluation but does not provide – opposed to most of the introduced tools – specific questionnaires or checklists etc. It does not include specific considerations for a particular sector. The author applied the framework in practice and met with severe resistance in the process, which they attribute, amongst others, to the generally low commitment to evaluation (J. M. Brinkerhoff, 2002, p. 229).

**Canadian Coalition for Global Health Research's Partnership Assessment Tool**

The Canadian Coalition for Global Health Research has designed a Partnership Assessment Tool, including actors from the Global South in the development process as authors and through local consultations. It covers the various stages of partnerships, from initiation to termination and is supposed to incite discussions among the partners and serve as a 'living document' that gets continuously updated during the whole partnership's duration. It is aimed at research partnerships, specifically in the health sector, but can be transferred to other sectors and initiatives (Afsana et al., 2009).

**ESTHER Alliance for Global Health Partnerships' EFFECt Tool**

The EFFECt Tool by the ESTHER Alliance for Global Health Partnerships focusses on evaluating how learning occurs and change is embedded by the members, and the participation of members of the Global North and Global South. It is designed to ignite discussions amongst the partners and limited to MSI-internal use (ESTHER Alliance for Global Health Partnerships, n. d.).

**Manual for Monitoring and Evaluating Education Partnerships**

The UNESCO, the World Economic Forum and Partnerships for education have jointly issued an extensive manual for evaluating MSIs in the field of education. It covers various stages of a partnership and provides comprehensive background information on MSIs and case studies for the users. It might, however be lengthy and quite complex to follow and is tailored for education initiatives (Marriott & Goyder, 2009).

### MSI Evaluation Tool

The MSI Evaluation Tool has been developed by MSI Integrity and the International Human Rights Clinic at Harvard Law School (2017). It is a comprehensive tool dedicated to evaluating the effectiveness of MSIs, including its conceptualization of necessary factors, for example their governance and accountability mechanisms. The MSI Evaluation Tool is designed for MSIs in the human rights sector and would require modifications for its application to MSIs in other fields. It was designed for both, an external use by stakeholders and an internal use by MSIs themselves or single members, although not all queried information might be publicly available. It has been developed based on research, global consultation, and practical pilot-testing (MSI Integrity & International Human Rights Clinic at Harvard Law School, 2017).

### OECD DAC Principles for Evaluation of Development Assistance

The DAC Principles for Evaluation of Development Assistance have originally been developed for aid agencies from OECD countries as a reference for evaluating their development assistance, which the OECD consider a "a co-operative partnership exercise between donors and recipients" (OECD, 1991, p. 4), and have been the most widely used criteria in international development (Chianca, 2008, p. 41). They can also serve as a guide for evaluating MSIs, and have been used by several initiatives queried by the OECD partnership survey (OECD, 2008, p. 12). Based on gathered learnings upon implementation in practice and scholarly claims, such as shifting the focus of criteria to the final beneficiaries' instead of donors' and governments' interest (Chianca, 2008), the OECD updated and improved their DAC Principles in an extensive consultation process in order to align it with the 2030 Agenda. The criteria can be employed to evaluate the process as well as the results of an MSI at any time of the initiative's duration. They are, however, very concise and might not cover all aspects or give detailed information on their application (OECD, 2019).

### The Partnering Initiative's Internal Prospective Partnership Assessment Tool

The Partnering Initiative developed a tool that can be used in the scoping phase prior to concluding a partnership agreement for evaluating risks, value, and points still to be resolved of a potential partnership as a basis for the decision-making. It is a very short and simple tool that can be employed only during the initiation phase of MSIs (The Partnering Initiative, n. d.).

### The Partnership Assessment Tool

The Partnership Assessment Tool commissioned by the British Office of the Deputy Prime Minister and developed by the Nuffield Institute for Health aims at providing a quick and cost-effective basis for evaluating the effectiveness of MSIs. It is meant as a developmental support for MSIs rather than a central tool for comprehensive evaluation. It is based on generic principled and can therefore be applied to a wide amplitude of contexts. The tool has been successfully applied to partnerships in the realm of health and social care and has particularly been readjusted for its suitability for multi-stakeholder initiatives (Hardy et al., 2003).

## 2.8 Summary

The literature research on multi-stakeholder initiatives and their evaluation has shown that it is a cross-cutting issue, which means that publications are spread across different disciplines and often focus on different and very specific aspects (cf. chapter 2.2). Another aspect that proved to be complex with regard to providing the theoretical background for the empirical research and the research questions of this work, is the fact that for many terms and concepts fundamental to the topic, there are varying and heterogeneous definitions, conceptualizations, and approaches, and no universally accepted standards exist (cf. for example chapter 2.6 ‘Effectiveness’ or chapter 2.7.1 ‘Evaluation’). Academic literature confirms research deficits regarding the evaluation of MSIs (e.g. Biermann et al., 2007; Pattberg & Widerberg, 2016; Stadtler, 2016). Concerning the systematic analysis of evaluation tools in particular, solely one publication appeared during my literature search, which reviews evaluation tools in different context (Tsou et al., 2015).

Based on the publication of numerous authors, two main realms of expectations and assessment criteria became evident, namely legitimacy and effectiveness (cf. Widerberg & Pattberg, 2015). Legitimacy comprises aspects regarding participation, discursive practice, transparency, and accountability (Bäckstrand, 2012; Dingwerth, 2007). Effectiveness can be distinguished into output, outcome and impact effectiveness (e.g. Jastram & Klingenberg, 2018; Wolf, 2010), referring to self-committed goals, changes in human behavior in the process, and ultimate contribution to solving the addressed collective action problem. While other names and categorizations also exist, these two conceptualizations of legitimacy and effectiveness encompass most content aspects, which I came across during my literature research, and provide a solid background for the data analysis and discussion in the empirical part of this work.

# 3 Methods

This chapter explains the methods I employ in my empirical research. The first three subchapters describe the research design (chapter 3.1), the method of data collection (chapter 3.2) and the method of data analysis (chapter 3.3), as well as their suitability for my research objective. Finally, chapter 3.4 gives further information on the actual execution of the empirical research, addressing both phases of data collection and data analysis.

## 3.1 Description of Research Design

As discussed in the theoretical background of this thesis, the knowledge and research on the evaluation of MSIs are still limited, and few studies have systematically examined the evaluation of collective action efforts (e.g. Biermann et al., 2007; Stadtler, 2016; Ven et al., 2017). Regarding the systematic investigation of evaluation tools, I only came across one publication in my literature search, which reviews evaluation tools in another context (Tsou et al., 2015). Also, as outlined in the theoretical part, the evaluation of MSIs is a complex topic at the intersection of different disciplines.
Against this background, I employ a qualitative comparative study for this research. Qualitative research designs are usually more "open to what is new in the material being studied" (Flick et al., 2004, chap. 1, Sect. 2), and are suitable when there is limited knowledge about the research subject and to comprehend complex issues. In comparative studies, several cases are examined with regard to certain extracts – in this case, specific evaluation questionnaires or indicators – in order to analyze them comparatively (Flick, 2004, chap. 4.1, sect. 2).

There is a wide variety of methods within the realm of qualitative research; in the following two chapters I specify the methods used for data collection and data analysis.

## 3.2 Method of Data Collection

For the data collection of this thesis, I apply document analysis.
Since document analysis is a somewhat controversial and not very common method of data collection in social sciences (e.g. Hoffmann, 2012, p. 395), I first briefly introduce document analysis and its role in empirical research, as well as the nature of documents as data, and then point out why it is the method of choice for my research.

Document analysis is assumed to be one of the oldest scientific tools and is of crucial importance in historic research (Schmidt, 2017, pp. 443–444). However, the status of document analysis is disputed, including whether it is considered a method of data collection or data analysis (e.g. Hoffmann, 2012, p. 395; Prior, 2008, p. 822). Various overviews of empirical or qualitative social research do not consider document analysis as a topic of its own. Those which do, mostly attribute it to the methods of data collection (Hoffmann, 2012, p. 399). Several authors (e.g. Glaser & Strauss, 1967, p. 163; Schmidt, 2017, p. 444), however, regard document analysis as a specific instrument of data acquisition analogous to e.g. interviews and surveys, and stress its significance in empirical social research.

Documents are commonly defined as data in the form of written text. They include, amongst others, official documents or publications such as handbooks, guides, brochures etc. (Salheiser, 2014, p. 813; Schmidt, 2017, p. 446).
Documents are 'natural' data, and their acquisition represents a non-reactive method of data collection. That means the data is preexistent and has not been created for research purposes. Therefore it has been generated without the researcher's participation or intervention (Hoffmann, 2012, p. 397; Salheiser, 2014, p. 813; Schmidt, 2017, p. 445).

Data collection through document analysis has two major advantages, which make this method best suited for my research objective. Firstly, it allows for access where interviews or participatory methods are not accessible for the researcher, like to large international organizations in the case of this thesis. Secondly, this method, being non-reactive, avoids influence of the research on the research subject. For example, interviewees' answers might be influenced by considerations of 'social desirability' (Schmidt, 2017, p. 445). With respect to this thesis, interviewing staff members of organizations about their own evaluation practices would have presumably entailed such influential effects.

For a small number of documents, like in this master thesis, the selection of the analysis sample is commonly conducted in a material-guided manner. The inclusion of individual documents is based on their relevance to the research question as well as considerations on their credibility, authenticity and representativeness (Salheiser, 2014, p. 822; Schmidt, 2017, p. 449). The selection criteria reflecting these factors as well as the selected analysis sample are specified in chapter 3.4.1.

## 3.3 Method of Data Analysis

For the data analysis of this thesis, I employ the method of qualitative content analysis according to Mayring.
The qualitative content analysis by Mayring seeks to make the process of text analysis describable and intersubjectively verifiable by formulating content-analytical rules. A category system is at the core of the method and serves as the main analytical instrument.
The categories as well as the content-analytical rules for their attribution to the text must be defined. The category system is usually determined in a theory-guided manner, either in the form of deductive category application or inductive category formation. Categories are then assigned to the text material by rule-guided interpretation. Subsequent to the category assignment, however, some aspects such as category frequency can also be analyzed in a quantitative way (Mayring, 2020, pp. 497–498; Mayring & Fenzl, 2014, pp. 544–545).
Mayring (2020, p. 500) themselves consider 'qualitatively oriented category-guided text analysis' a more appropriate denomination of the analysis method.

Of the three basic techniques of qualitative content analysis proposed by Mayring (2014, 2015), I employ 'structuring' for the analysis of the collected data. Structuring content analysis makes it possible to obtain a cross section of the document material and filter out specific aspects along predefined criteria. For that purpose, the entry data gets systematized according to deductively derived, predetermined categories (Mayring, 2014, p. 64, 2020, p. 497). The single steps of the deductive category assignment process are depicted in Figure 8.

On a general level, content-analytical methods are suitable for the analysis of small numbers of documents, like in this thesis (Salheiser, 2014, p. 815). In particular, Mayring's structuring technique is very appropriate for my research purpose: The preliminary theory-based deduction of categories permits to establish a rough analytical frame for the inhomogeneous entry data, which consists of evaluation tools varying strongly in scope, target audience, sector, partnering phase etc. (cf. Table 3). Although the research on MSI evaluation is limited (cf. chapter 3.1), the theoretical background provides a sufficient basis for delineating an initial category system. At the same time, the technique allows to integrate the additional knowledge gained during the preliminary run-through of the data by inductively refining the category system in a next step, and re-running the coding based on the revised categorization (cf. Figure 8).

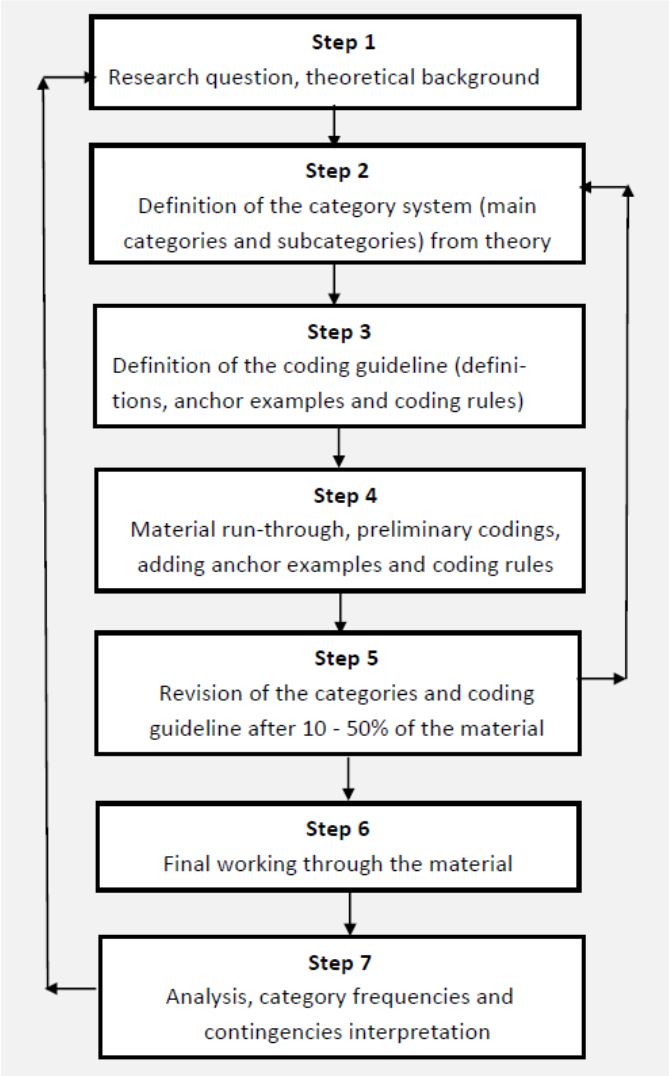


*Figure 8*. Process model for the deductive category assignment according to Mayring (Mayring, 2014, p. 96).

## 3.4 Execution of the Research

In this chapter, I describe the practical execution of the empirical research as well as challenges, limitations, and recommendations I encountered in the process. Since the selection of the research sample and the actual data collection are closely linked for the method of document analysis, they are jointly outlined in subchapter 3.4.1. In subchapter 3.4.2, I provide details on the data analysis phase.

### 3.4.1 Selection of Research Object & Execution of Data Collection

Since, firstly, the denomination of collective action efforts varies widely (cf. chapter 2.1.2) and, secondly, assessment tools are mainly published by institutions, not academic researchers, I was unable to perform a search on scientific publication platforms by key words. Also, an initial

broad internet search didn’t turn up many evaluation tools. Instead, I followed all cross-references to assessment tools I came across while researching the theoretical background.
In a next step, I established a set of criteria by which I selected the tools to be included in the analysis. The criteria reflect the considerations for selection mentioned in chapter 3.2, and are listed below. A final sample of eleven evaluation tools met the selection criteria and is specified in Table 3.

The criteria for selection are:

- Directed at multi-stakeholder initiatives
- Includes a set of specific questions or indicators for evaluation
- Published in a scientific journal or by an official institution
- Available in English
- Clearly discernable date of publication
- Non-modifiable data format (PDF files only)

As the research objective is directed at evaluation tools for MSIs, the analysis does not include general evaluation guidelines, manuals, and frameworks, nor documents including only exemplary questions or indicators. Also, tools designed for single potential members to assess whether or not to engage in a MSI do not form part of the research object.

A shortcoming of this selection process, however, is the lack of fixed search parameters, and therefore also the lack of an exhaustive sample list within those parameters. Consequently, the sample in my research is not objectively replicable, and other researchers presumably would arrive at different data samples.

| ID | Issuing organisation | Author(s) | Title of publication / tool | Year of publication | Analyzed part within publication | Target sector / MSIs | Partnering phase(s) |
|---|---|---|---|---|---|---|---|
| T1 | - | Brinkerhoff, J. M. | Assessing and improving partnership relationships and outcomes: a proposed framework | 2002 | Summary of proposed assessment targets and methods, pp. 221-223 | - | Inception, Implementation |
| T2 | Canadian Coalition for Global Health Research | Afsana, K., Habte, D., Hatfield, J., Murphy, J., & Neufeld, V. | Partnership Assessment Toolkit | 2009 | Partnership Assessment Tool, pp. 15-24 | Health research | Inception, implementation, dissemination, termination |
| T3 | Center for the Advancement of Collaborative Strategies in Health | | Partnership Self-Assessment Tool | 2002 | Questionnaire, pp. 2-15 | Health | Implementation |
| T4 | Compassion Capital Fund National Resource Center | | Partnerships: Frameworks for Working Together | 2010 | Questions for evaluation and monitoring, pp. 27-28 | Social community services | Implementation |
| T5 | Office of the Deputy Prime Minister of the United Kingdom; Nuffield Institute of Health | Hardy, B., Hudson, B., & Waddington, E. | Assessing Strategic Partnership: The Partnership Assessment Tool | 2003 | Undertaking the partnership assessment, pp. 15, 19, 23, 27, 31, 35, 39 | Public service | Inception, Implementation |
| T6 | National University of Ireland Galway | Mahmood, S.; Morreale, S.; Barry, M. | Developing a Checklist for Intersectoral Partnerships for Health Promotion | 2015 | Partnership Checklist: Tool for Assessing Intersectoral Partnership Functioning, pp. 12-14 | Health | Inception, Implementation |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| T7 | UNESCO IIEP; World Economic Forum; Partnerships for education | Marriott, N.; Goyder, H. | Manual for monitoring and evaluating education partnerships | 2009 | Tool 5. Checklist for evaluating a partnership's Strengths and weaknesses, pp.109-112 | Education | Implementation |
| T8 | MSI Integrity; International Human Rights Clinic at Harvard Law School | | MSI Evaluation Tool for the Evaluation of Multi-Stakeholder Initiatives | 2017 | pp. 12-59 | Human rights, standard-setting MSIs | Implementation |
| T9 | Nonprofit Finance Fund; Center for Health Care Strategies; Alliance for Strong Families and Communities | | Partnership Assessment Tool for Health | 2017 | pp. 3-10 | Health | Inception, Implementation |
| T10 | The Partnering Initiative | | Partnership health check | 2018 | pp. 1-2 | - | Implementation |
| T11 | Victorian Health Promotion Foundation | | The partnerships analysis tool. A resource for establishing, developing, and maintaining partnerships for health promotion | 2016 | Checklist, pp. 6-7 | Health | Inception, Implementation |

*Table 3*. Selected data sample for analysis.

### 3.4.2 Execution of Data Analysis

This chapter delineates the practical execution of the data analysis, including the software used, the initial settings for the analysis, the development of the category system, and the export data resulting from the analysis.

#### Software QCAmap

For conducting the data analysis, I employed the software program QCAmap[5], which has been specifically developed for qualitative content analysis by Mayring et al. (n. d.) and comprises the whole analysis process (Mayring, 2020, p. 502).

While it is an advantage that the software program has been specifically tailored to the employed analysis method, its coding and exporting functions are not (yet) very user-friendly. For example, the preliminary coding cannot be centrally reset after the pilot phase but needs to be deleted separately for each document. Also, the analysis and export options are limited, and the exported category system lacks any formatting and had to be completely re-compiled for display in this thesis (cf. Appendix A).

#### Data Entry & Analysis Settings

For the upload to the software program, I was required to convert the documents into a DOCX format, manually reduce them to the parts to be analyzed (cf. Table 3, column 6) and, where necessary, correct errors that emerged during the import process.

Subsequently, I specified the units for the analysis. Since the whole entry data was analyzed following the same category system, I determined the entire material as the recording unit, i.e. the portion of text that is compared to the same category system (Mayring, 2019, chap. 3.4). As coding unit, which represents the smallest portion of material that can be coded and therefore the sensitivity of the analysis (Mayring, 2015, p. 61, 2019, chap. 3.4), I defined one complete sentence or question. However, questions with subsequent follow-up questions such as “If yes, which one?” were coded as one unit. The context unit, being the whole material that exists on one case (Mayring, 2014, p. 100), was the entire respective evaluation tool.

I chose to allow for multiple categorizations in the coding, as on the one hand, some categories are closely interrelated and one question might relate to more than one category, and on the

[5] The software program is freely accessible upon registration at www.qcamap.org.

other hand, difficulties in assigning a single category might provide valuable information on the clearness of evaluation questions (cf. chapter 4.2).

**Category System**

As described in chapter 3.3, I applied the technique of deductive category assignment for the qualitative content analysis. Based on scientific literature, I initially established the categorization displayed in the second column of Table 4. After the pilot phase and preliminary run-through of five evaluation questionnaires (corresponding to 45 % of the material in number of tools and 25 % in number of pages), I revised the categories and coding guideline (cf. Figure 8) and inductively added the subcategories listed in the third column of Table 4. The detailed coding guideline including definitions, anchor examples and coding rules is attached in Appendix A.

| No. | Deductively derived categories | Inductively derived subcategories |
|---|---|---|
| 1 | Participation & Inclusiveness | internal<br>external |
| 2 | Transparency | internal<br>external |
| 3 | Accountability & Control Mechanisms | internal<br>external |
| 4 | Responsiveness | - |
| 5 | Discourse & Deliberative Practice | - |
| 6 | Institutional Effectiveness | Vision & Purpose<br>Roles & Responsibilities<br>Resources<br>Communication<br>Ethics & Work Culture<br>Commitment<br>Benefits<br>Organizational Structures & Change<br>Organizational Form<br>External Support<br>Termination<br>Other |
| 7 | Output Effectiveness | - |
| 8 | Outcome Effectiveness | - |
| 9 | Impact Effectiveness | - |

*Table 4*. Category system for data analysis.

**Export of Analysis Results**

After completing the categorization, the following files can be downloaded from the software program:

- Coding guideline (see previous paragraph and Appendix A)
- Coded passages
- Document statistics

The file of coded passages lists for each category all text units that have been assigned to it, including the entry document they originate from. The file's content is summarized in the first section of chapter 4.1. The downloaded document statistics indicate the number of text units assigned to each category per analyzed document; the results are depicted in the second part of chapter 4.1.

Another analysis feature which would be insightful but is unfortunately not included in the software, are statistical results indicating the frequency of multiple categorizations and showing between which categories they occurred. A drawback regarding the analysis results is also that the page or line number of the coded text unit is not included in the export file, and therefore the entry documents must be reviewed manually for each passage when displaying the results.

The empirical results of the analysis are detailed in chapter 4. In chapter 5, I provide a discussion of these results in relation to the theoretical background (chapter 2) and the research questions (cf. chapter 1.2) of this thesis.

# 4 Empirical Results

In this chapter, I present the results of my empirical data collection and analysis. The first subchapter details which evaluation aspects the analyzed evaluation tools comprehend. The second subchapter describes further results of the analysis, obtained on the basis of categorization overlaps, and regarding the specificity of the analyzed questionnaires for particular target sectors.

## 4.1 Evaluation Aspects Covered by Current Evaluation Tools

In this chapter, I describe which aspects of evaluation are comprised by the evaluation questionnaires. First, I outline the evaluation subjects included in the analyzed tools. In the second part, I specify the quantitative category frequencies resulting from the analysis.

One issue the analyzed evaluation tools address, is the general aptitude of the MSI to contribute to the respective collective action problem. On the one hand, it includes the assessment whether a MSI approach is the most adequate way to address the problem (T6, p. 12), and whether there is a perceived need for collaboration in terms of complementary skills, capacity, and common interest (T6, p. 12; T11, p. 6). On the other hand, it comprises evaluating how well the MSI is contributing and responding to the needs of the beneficiaries and/or community. This query is either addressed in a general way, such as “By working together, how well are these partners able to respond to the needs and problems of the community?” (T3, p. 2), or specifically regarding the sector and beneficiaries the evaluation tool is designed for, such as “Does the MSI comprehensively address the human rights issues that prompted its formation?” (T8, p. 13), or “How is the partnership designed to assist both learners and teachers improve educational outcomes?” (T7, p. 110). In terms of development over time and continuous monitoring and revision, T4 (p. 27) and T10 (p. 2) question whether the MSI still meets its original need for existence and remains relevant to the context, and T7 (p. 110) explores if the MSI needs to take actions to improve its relevance.

Another area of evaluation is the participatory and inclusive quality and transparency of the MSI. On the one hand, this comprises the participation, inclusiveness, and transparency within the MSI, with respect to membership, processes, and organizational bodies. On the other hand, it assesses the involvement of external stakeholders and the transparency towards external actors and the public.

With regard to the composition and membership of the MSI, evaluation tools inquire which groups form part of the MSI (T8, p. 21), whether all relevant stakeholders are represented (T4, p. 27; T6, p. 12), and whether the variety of members allows for a comprehensive understanding of the problem at hand (T6, p. 12; T11, p. 6). T8 (p. 23) additionally examines the geographical representation within the stakeholder groups, and T3 (p. 6) addresses the diversity of people and organizations forming part of the MSI. Further evaluation questions deal with the involvement of MSI members in different stages of the MSI process (T2, p. 17; T7, p. 109), as well as in different bodies of the MSI (T8, pp. 24-26). T8 (p. 18) explores barriers for participation, such as membership fees or working language. T11 (p. 7) includes the possibility of reviewing and changing the range of members.

Regarding the participatory and inclusive culture of MSI-internal deliberations and decision-making processes, evaluation questions range from structures and mechanisms to members' subjective perceptions. Structural aspects include the frequency, logistics and language of meetings (T3, p. 8; T8, p. 26; T10, p. 2), the decision-making procedures (T8, p. 26), as well as structures and strategies for ensuring the expression of alternative views and the resolution of conflicts (T1, p. 221; T6, p. 13; T11, p. 7). The evaluation of members' perceptions comprises their influence on decisions and activities (T3, pp. 13-14; T7, p. 109), the consideration of different views (T1, p. 221), the openness to critical feedback and challenging assumptions (T1, p. 222; T6, p. 13; T7, p. 109; T8, p. 16), and the ability to share power (T1, p. 221; T6, p. 13).

Regarding internal transparency, evaluation questionnaires mainly assess the documentation and circulation of information on decisions, results, and findings, as well as the corresponding structures and processes (T1, p. 222; T2, p. 24; T5, p. 35; T6, p. 14; T9, p. 9). Also covered is the transparent communication of resources and their management (T6, p. 13; T7, p. 109; T8, p. 18), of the appointment or election processes to MSI bodies (T8, p. 24), and of members' expectations, motivations, and needs (T10, p. 2). T3 (p. 7) and T10 (p. 2) relate to the provision of documentation as a basis for decision-making, and T1 (p. 222) takes into consideration the "timely response to information requests".

As for the participation of external stakeholders, analogous to the internal participation, evaluation tools address the question which stakeholder groups have been engaged in the process, such as community members, government bodies, academic stakeholders, funders etc. (T2, p. 21; T4, p. 27; T6, p. 14; T8, pp. 29-36, 55-56; T9, p. 4). T8 (pp. 30-33, 35, 49-50, 55-56) investigates input and feedback processes, to whom they are open, and at which stages of the MSI they occur. T8 (pp. 17-18, 37) also includes the identification of languages spoken by the stakeholders, as well as the geographic location of contact points and meetings, and meeting

frequency. T9 (p. 6) encompasses questions on how the involvement and participation of the target population and communities can be improved.

Regarding the MSI's transparency towards external stakeholders and the public, questionnaires explore whether the MSI's purpose, goals, and planned actions regarding the collective action problem at hand, are clear to the public (T3, p. 3; T4, pp. 27-28; T10, p. 2). Other evaluation aspects include the public availability of information about the MSI and its members, bodies, and decisions (T4, p. 27; T8, pp. 17-18, 26, 27, 47, 54). T8 (p. 59) furthermore covers the publication of internal reviews. T8 (pp. 19-20) and T9 (p. 7) investigate which details about expenses are shared with which stakeholder groups. Regarding the MSI's results, T2 (p. 17) explores the existence of a dissemination plan for research results, and T8 (p. 49) and T9 (p. 10) examine the publication of outcomes and recommendations. T8 (pp. 14, 18, 27, 28, 35, 48) includes a very detailed set of questions regarding the identification of languages spoken by the affected stakeholders and the provision of information in these languages. T8 is tailored to standard-setting MSIs and specifically addresses publicly available information on the standards (p. 14), as well as on the actors committed to the standards and their businesses (pp. 40-41). It also details the reporting duties of committed actors (T8, pp. 34-35, 40), for example if they are obliged to report directly to the public (p. 34). T8 additionally inquires whether the MSI or actors committed to the standards offer programs providing information on the MSI, the standards or the addressed collective action problem, in which geographical region they offer them and for which target group(s) (pp. 41-43). Furthermore, T8 comprises evaluation questions regarding the transparency of the grievance system, dealing with breaches of internal governance rules, and of filed complaints (pp. 27, 45, 48).

Evaluation tools also assess the internal structures, arrangements, and culture of the MSI, relating for example to the purpose of the MSI, the distribution of work and resources, the culture of working together and the commitment to the MSI, as well as to the organizational and communication structures.

A number of questions are directed at the shared mission, vision, aims, and objectives of the MSI, and whether they are mutually agreed upon and clear to all members (T1, pp. 221-223; T2, p. 15; T3, p. 2; T4, p. 27; T5, p. 19; T6, p. 12; T8, p. 12; T10, p. 2; T11, p. 6). T2 (p. 19) and T4 (p. 27) further inquire about their possible evolution during implementation, and T6 (p. 13) investigates the leadership's process towards consensually reaching them. T11 (p.7) asks whether differences in organizational priorities are dealt with.

With regard to the ethical practice of the MSI, evaluation tools examine the ethical conduct within the MSI and its evolution during implementation (T2, pp. 16, 19, 23). T5 (p. 12) assesses whether the MSI "is planned to be culturally appropriate to the public/community/local setting", and T8 (p. 30) specifically questions whether evaluators have experience or training in cultural sensitivity. In the context of research MSIs, T2 (p. 16) determines whether all necessary ethical clearances have been obtained. In the context of both, ethical practice and work culture, T7 (p. 109) inquires whether the MSI has determined a shared code of conduct for its operations, and whether the MSI has followed it in the perception of its members. Concerning the working culture within the MSI, evaluation questionnaires explore whether the members work well together and have good relations (T3 p. 14; T4, pp. 27-28; T6, p. 12; T11, p. 6), and address the trust, respect, and openness between them (T1, pp. 221-222; T3, p. 5; T4, p. 28; T5, p. 27; T6, pp. 12-13; T10. p. 2). T5 (p. 27) investigates whether the trust levels within the MSI "are high enough to encourage significant risk-taking". Further questions aim at examining whether the MSI in its design and work recognizes and appreciates the contributions of each member as well as joint achievements (T1, p. 222; T3, p. 13; T5, p. 27; T11, p. 7), and if members feel empowered, enabled, and motivated (T3, p. 4; T10, p. 2). T2 (p. 19) explores mentorship, and T3 (p. 8) assesses how well the MSI provides orientation to new members joining the initiative. Other evaluation aspects cover the mutual understanding of drivers and motivations of each member (T1, p. 222; T10. P. 2) and the reliability of members (T1, p. 221; T10, p. 2). T1 (pp. 221-223) explores the culture and underlying core values of the member organizations, their compatibility, and the development of a MSI organizational culture.

Regarding the commitment to the MSI, evaluation questions are directed at the commitment of members and individuals to the MSI objectives and to good collaboration (T4, p. 27; T5, p. 23; T6, p. 12; T11, p. 6), the robustness of the commitment (T5, p. 23), as well as its medium-term persistence (T11, p. 7). T1 (p. 222) and T4 (p. 27) examine the readiness of the members to make adjustments in favor of the MSI. T9 (p. 4) investigates whether support is lacking from individuals of the member organizations, and how the issue has been responded to. Further aspects of commitment addressed in the questionnaires are the support from high level management of the member organizations (T1, p. 221; T5, p. 23; T9, p. 4; T10, p. 1; T11, p. 6), and the encouragement of collaborative action in the member organizations and the MSI (T5, p. 23; T11, p. 7).

Another set of questions concerns the roles and responsibilities within the MSI and whether they are clearly defined and understood (T2, p. 16; T4, p. 27; T5, p. 31; T6, p. 13; T11, p. 6), as well as the satisfaction of members with their distribution (T2, p. 23; T3, p. 14). T9 (p. 6)

specifically addresses the staffing model of the MSI. Evaluation questions also take into consideration the compliance of members with the agreed roles and responsibilities (T1, p. 223; T10, p. 2). T6 (p. 13) and T11 (p. 6) inquire whether some roles "cross the traditional boundaries that exist between agencies or divisions in the partnership". T5 (p. 15) investigates the partial interdependence and partial independence of members' activities in achieving the MSI's objectives. T2 (pp. 18, 21, 23), which is designed for research MSIs, includes questions about agreements regarding the ownership of products and intellectual properties generated by the MSI.

Regarding the resources of the MSI, questionnaires address the adequacy of the available financial and non-financial resources, such as infrastructure, data, skills, and expertise (T3, pp. 10-11; T6, p. 13; T8, p. 19; T10, p. 1). Evaluation tools also investigate the resources contributed to the MSI by each member, whether they are jointly agreed upon and appreciated, and whether members are satisfied with their allocation (T1, p. 222; T2, pp. 16, 19, 22-23; T5, p. 31; T9, pp. 3, 7; T11, p. 7). T9 (pp. 3, 6) expands on the necessary skills and expertise, and whether the current members cover them, or if additional skills and expertise are required. T3 (p. 6) examines the efficient use of resources, and T8 (p. 19) specifically questions whether the ratio of expenditure on implementation to total expenditure exceeds 33%. On the issue of financial resources, evaluation questions address financing and funding opportunities (T8, pp. 18-19; T9, p. 7), their sustainability (T9, p. 7; T11, p. 7), and the efficiency of their management (T3, p. 7; T10, p. 2). T9 (p. 7) investigates in detail the reliability of the MSI's sources of revenue, the level of current expenses and investment needs, and the prediction of their future development, as well as the management of unanticipated expenses.

As for the benefits of participating in the MSI, the analyzed tools evaluate what organizations and individuals expect to gain from engaging in the MSI and whether these reasons and motivations are jointly understood and accepted (T2, p. 15; T4, p. 27; T5, p. 19; T9, p. 8; T10, p. 2; T11, p. 6). T5 (p. 13) additionally questions whether the extent to which members participate in the MSI involuntarily, due to pressure or mandate, are acknowledged. Another set of questions surveys what benefits individual member organizations have gained through their engagement in the MSI, such as enhanced performance in servicing constituencies, increased public profile, newly acquired skills and knowledge, additional financing, amplified network, or improved ability to make a contribution to society (T1, pp. 222-223; T3, pp. 12-13; T7, p. 111; T10, p. 2; T11, p. 6). T3 (p. 13) also inquires the drawbacks experienced by the member organizations, such as redirection of resources away from other operations, conflict of interests, deteriorated image due to the affiliation with other members of the MSI, or insufficient credit for contributions to the MSI. T9 (p. 8) determines the financial risk each member takes by

participating in the MSI and the expected effect on each member's bottom line. Further questions investigate the balance of benefits and costs or drawbacks for the members (T3, p. 14; T11, p. 6), and the fairness of benefit distribution (T1, p. 222; T5, p. 27).

With respect to the organizational structure of the MSI, evaluation tools examine the MSI's governance structure and possible changes to it during implementation (T2, pp. 15, 19, T8, p. 57). T8 (pp. 24, 26) investigates the existence, composition, and competences of bodies for different decision-making, implementation, and reviewing functions. Further questions cover the MSI's leadership and its capacities (T3, pp. 4-6; T6, p. 13), as well as how changes in the leadership would affect the MSI (T9, p. 4). T1 (p. 221) and T2 (pp. 15, 19) also assess whether institutional agreements for the MSI exist, and their level of formality. T9 (p. 9) specifically examines the MSI's systems for data collection, revision, and analysis, and T8 (pp. 44-47 addresses in detail the grievance mechanisms and evaluation process of complaints. T5 (p. 31) and T11 (pp. 6, 7) assess whether operational and administrative arrangements are simple and task-oriented, and whether processes that apply to all members have been standardized. Another aspect of evaluation includes the structure and frequency of meetings, and their documentation routine (T1, p. 221; T2, p. 19; T10, p. 2). Regarding monitoring and revision systems and the MSI's approach to change, questionnaires on the one hand explore the monitoring and revision processes of the MSI objectives, and their possible evolution in the light of challenges and changes (T1, pp. 221, 222; T2, p. 16; T4, p. 27; T5, p. 35; T7, p. 110; T9, p. 3). On the other hand, they assess the arrangements for monitoring, evaluation, and improvement of the MSI itself (T1, p. 223; T2, pp. 15, 19; T4, pp. 27, 28; T5, p. 35; T6, p. 14; T8, pp. 24, 56-57, 59; T9, p. 9, T10, p. 2, T11, p. 7). Further evaluation questions address the MSI's and its members' ability to respond to changes in the environment or target population demands and to unanticipated situations (T1, p. 221; T9, pp. 5, 9), to find new and creative solutions (T1, p. 221; T3, p. 2), and to react to the need for corrective measures (T1, p. 221), as well as the structures and processes supporting those abilities. T9 (p. 9) and T10 (p. 2) further assess whether the MSI makes adjustments based on prior experiences and collected data.

Evaluation questions concerning communication structures and activities relate to both, communication within the MSI and undertakings of external communication, marketing, and representation. Regarding internal communication, questionnaires explore if the processes, lines, and tasks of communication among the members are clear and effective (T1, p. 222; T2, pp. 17, 23; T3, p. 7; T6, p. 13; T10, p. 2; T11, pp. 6-7). Furthermore, they inquire about opportunities for formal and informal communication and face-to-face communication among members (T4, p. 28; T6, p. 13; T11, p. 7), the encouragement of networking (T5, p. 23), and the promotion of

the MSI within the members' organizations (T1, p. 221; T11, p. 6). With respect to external publication and representation activities, evaluation questions examine the communication strategy and dissemination plan in terms of timeline and media of distribution (T2, pp. 18, 21; T4, p. 27; T6, p. 14), as well as the coordination and channels of communication with stakeholders outside the MSI (T3, p. 7; T5, p. 35; T8, p. 42; T9, pp. 4, 10). T2 (pp. 18, 21), which is directed at research MSIs, also includes the representation at conferences and the scientific publications of the MSI (T2, pp. 18, 21). T1 specifically addresses the name recognition of the MSI (p. 223) as well as regulations determining who can represent the MSI (p. 222).

Concerning external support for the MSI, T9 (p. 4) inquires in general which external stakeholders support the MSI, and how their support can be enhanced. Other evaluation questions more specifically investigate the support for the MSI and the MSI objectives from leaders, politicians, and decision-makers (T3, p. 9; T4, p. 28; T6, p. 26; T7, p. 110), and from the community (T3, p. 3). Also, connections and strategic alliances with other organizations and actors are included in the assessment (T1, p. 223; T3, p. 9; T6, p. 27). T8 (p. 40), a tool designed for standard-setting MSIs, further addresses whether the MSI promotes business relations among actors committed to the standards, and whether it prohibits business relations with industry actors who do not comply with the MSI standards.

Evaluation questionnaires further cover the identification of supporting factors and barriers to MSI success (T1, p. 223; T5, p. 13; T6, p. 26) and of MSI strengths and concerns (T2, p. 20; T7, pp. 111-112). Another investigated aspect is whether the MSI takes into consideration other MSIs operating in the respective industry (T8, p. 54), as well as the political, economic, social, and cultural context affecting its work (T6, p. 26; T9, p. 9).

Regarding accountability mechanisms established by the MSI, evaluation questions examine in general whether the formal structures are designed to support accountability (T6, pp. 13, 14; T11, p. 6), and whether there are "clear lines of accountability for the performance of the partnership as a whole" (T5, p. 31). T1 (p. 222) specifically investigates the reporting, performance information, and financial controls among members. T8 (pp. 15-16, 26-27, 29-35, 43-47, 49-50), which is tailored to the evaluation of MSIs setting human rights standards, in-depth assesses the requirements and accountability of committed industry actors with respect to their compliance with the MSI standards. This includes the timeline for compliance, reporting duties, evaluation procedures and obligations, and the procedures and requirements for managing grievances concerning breaches of internal governance rules, MSI standards, or other human rights standards, as well as possible sanctions.

With respect to external accountability, T8 (p. 50) includes the question whether the MSI "enable[s] or encourage[s] public enforcement of its standards in national and/or international processes".

Regarding the results of the MSI, evaluation tools inquire whether the MSI has defined goals, and whether they are known and agreed upon by the members (T1, p. 221; T2, p. 15; T3, p. 2; T4, p. 27; T5, p. 19 T6, p. 12; T10, p. 1; T11, p. 6). Further questions assess if the goals are realistic and achievable (T6, p. 12; T7, p. 110; T9, p. 3), and how they are measured and monitored (T4, p. 27; T5, p. 35; T7, p. 110; T9, p. 9). T9 (pp. 9-10) additionally specifies the data, systems, and people necessary to understand the outputs and the progress towards them. Another set of questions investigates the progress in attaining the projected output goals (T1, pp. 222-223; T2, p. 19; T4, p. 27; T5, p. 39; T9, p. 3; T10, p. 2), and how often the progress and the goals themselves and are reviewed and discussed (T1, p. 221; T4, p. 27; T7, p. 110; T9, pp. 3, 10). T9 (p. 3) also examines which aims exceed the scope of the MSI and are not targeted. T4 (p. 27) and T7 (p. 110) furthermore explore how the MSI's work links to local and national policies. T8 comprises a detailed set of questions for standard-setting MSIs with respect to the standards and the actors committed to them. It includes which industries the standards are designed for (p. 12), for which geographical regions they apply (p. 12), how they relate to international law (pp. 14-16) and how actors are incentivized to follow them (p. 28). T8 also inquires if the MSI provides recommendations for the actors committing to the standards (pp. 48-49), and if it offers learning programs about the collective action effort – in this case human rights – and the MSI standards, and for which stakeholder groups (pp. 35-37, 41).

On the level of awareness and capacity building and behavioral changes, T2 (pp. 16, 19) takes into consideration capacity building goals and skills development, and T8 (p. 58) includes the question whether and regarding which aspects the MSI assesses the awareness of affected actors.

With regard to the effects on the beneficiaries and the collective action problem, one main evaluation question is whether the MSI adds value and makes a contribution to the community (T3, p. 13; T6, p. 14; T11, p. 7). T8 (pp. 57-58) examines the review process of the actual and potential impact on local communities and of the impact on other affected populations, specifically also addressing whether "the MSI has a discriminatory impact on any population groups or subgroups" (p. 58). T7 (p. 110) and T8 (p. 13) specifically assess the impact for their target sector – education and human rights, respectively – and T7 includes concrete indicators as basis for the assessment. T9 (p. 9) inquires which data the MSI collects to understand its impact, in

this case referring to clinical or social impact of the MSI. T9 (p. 5) further explores what additional services would be needed by the target population, and whether they lie within the scope of the MSI objectives.

Evaluation tools also address preparatory arrangements for the closing phase of the MSI. Questionnaires assess whether the MSI has an exit strategy and a closing plan in place (T2, p. 23; T6, p. 28; T7, p. 111). T2 (p. 23) subsequently investigates the plans for redeployment of staff and resources and for the management of MSI property after the finalization of the MSI. Concerning possible follow-ups, T2 (p. 24) examines if the members have scheduled meetings to explore future possibilities, directions, and collaboration opportunities, and T7 (p. 111) questions whether there have been deliberations on future funding possibilities for collective action activities after the MSI's termination.

**Category Frequency**

Qualitative content analysis according to Mayring also permits to analyze some quantitative aspects like category frequency (cf. chapter 3.3). Table 5 displays the number of coded text passages per (sub)category and analyzed document. In the last column, I additionally calculated the respective category frequency as a percentage of the total number of coded passages.

| **Category Frequency** | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| No. | Category / Subcategory | Document ID | | | | | | | | | | | SUM [Sub-cat.] | SUM [Category] | % of total |
| | | T1 | T2 | T3 | T4 | T5 | T6 | T7 | T8 | T9 | T10 | T11 | | | |
| 1 | Participation & Inclusiveness | | | | | | | | | | | | | **103** | 9,6 |
| 1.1 | internal | 2 | 8 | 3 | 4 | 0 | 2 | 1 | 18 | 1 | 0 | 3 | **42** | | 3,9 |
| 1.2 | external | 0 | 9 | 3 | 2 | 0 | 1 | 4 | 34 | 8 | 0 | 0 | **61** | | 5,7 |
| 2 | Transparency | | | | | | | | | | | | | **87** | 8,1 |
| 2.1 | internal | 4 | 1 | 1 | 1 | 1 | 3 | 1 | 4 | 3 | 3 | 0 | **22** | | 2,1 |
| 2.2 | external | 0 | 2 | 1 | 4 | 0 | 0 | 0 | 53 | 4 | 1 | 0 | **65** | | 6,1 |
| 3 | Accountability & Control Mechanisms | | | | | | | | | | | | | **121** | 11,3 |
| 3.1 | internal | 3 | 0 | 0 | 0 | 1 | 2 | 0 | 113 | 0 | 0 | 1 | **120** | | 11,2 |
| 3.2 | external | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | **1** | | 0,1 |
| 4 | Responsiveness | 0 | 0 | 4 | 2 | 0 | 0 | 5 | 2 | 2 | 1 | 1 | | **17** | 1,6 |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 | Discourse & Deliberative Practice | 13 | 4 | 11 | 0 | 0 | 8 | 6 | 22 | 3 | 5 | 5 | | **77** | 7,2 |
| 6 | Institutional Effectiveness | | | | | | | | | | | | | **491** | 46,0 |
| 6.1 | Vision & Purpose | 11 | 4 | 3 | 9 | 3 | 3 | 0 | 2 | 0 | 2 | 3 | **40** | | 3,7 |
| 6.2 | Roles & Responsibilities | 4 | 9 | 3 | 3 | 4 | 4 | 0 | 1 | 5 | 1 | 5 | **39** | | 3,7 |
| 6.3 | Resources | 3 | 6 | 10 | 0 | 2 | 5 | 0 | 12 | 21 | 7 | 5 | **71** | | 6,6 |
| 6.4 | Communication | 4 | 12 | 3 | 5 | 2 | 7 | 0 | 4 | 3 | 3 | 5 | **48** | | 4,5 |
| 6.5 | Ethics & Work Culture | 16 | 8 | 8 | 4 | 5 | 6 | 2 | 2 | 0 | 5 | 2 | **58** | | 5,4 |
| 6.6 | Commitment | 5 | 0 | 1 | 2 | 4 | 3 | 0 | 0 | 2 | 2 | 6 | **25** | | 2,3 |
| 6.7 | Benefits | 11 | 2 | 19 | 3 | 3 | 2 | 1 | 0 | 5 | 2 | 6 | **54** | | 5,1 |
| 6.8 | Organizational Structures & Change | 15 | 15 | 7 | 6 | 4 | 9 | 2 | 19 | 14 | 6 | 5 | **102** | | 9,6 |
| 6.9 | Organizational Form | 2 | 0 | 0 | 0 | 0 | 3 | 0 | 1 | 1 | 0 | 2 | **9** | | 0,8 |
| 6.10 | External Support | 1 | 0 | 2 | 2 | 0 | 3 | 2 | 3 | 6 | 0 | 0 | **19** | | 1,8 |
| 6.11 | Termination | 0 | 5 | 0 | 0 | 0 | 1 | 2 | 0 | 0 | 0 | 0 | **8** | | 0,7 |
| 6.12 | Other | 2 | 2 | 0 | 0 | 4 | 3 | 2 | 4 | 1 | 0 | 0 | **18** | | 1,7 |
| 7 | Output Effectiveness | 9 | 5 | 3 | 8 | 8 | 1 | 4 | 60 | 14 | 2 | 2 | | **116** | 10,9 |
| 8 | Outcome Effectiveness | 1 | 4 | 0 | 0 | 0 | 0 | 1 | 11 | 3 | 0 | 1 | | **21** | 2,0 |
| 9 | Impact Effectiveness | 4 | 0 | 3 | 3 | 0 | 1 | 3 | 11 | 8 | 0 | 2 | | **35** | 3,3 |
| | SUM [per document] | 110 | 96 | 85 | 58 | 41 | 67 | 36 | 377 | 104 | 40 | 54 | | **1068** | |

*Table 5*. Category frequencies of the analyzed data.

As can be seen in Table 5, the frequencies of the categories relating to democratic legitimacy (cf. chapter 2.5.6), ‘Participation & Inclusiveness’ (9,6 %), ‘Transparency’ (8,1 %), ‘Accountability & Control Mechanisms’ (11,3 %) and ‘Discourse & Deliberative Practice’ (7,2 %) lie in a similar range. Only the category ‘Responsiveness’ has a considerably lower frequency, with 17 coded passages, i.e. 1,6 % of the total. Within the category ‘Participation & Inclusiveness’, 42 units refer to participation and inclusiveness within the MSI and 61 units to external stakeholders. Within the category ‘Transparency’, one quarter corresponds to internal, and three quarters to external aspects of transparency. The discrepancy within the category of

‘Accountability & Control Mechanisms’ is substantial, with only 1 out of 121 coded units referring to external accountability, and the remaining 120 to accountability and control mechanism established by the MSI.
Regarding the categories referring to the effectiveness of the MSI, ‘Institutional Effectiveness’ accounts for 46,0 % of the total coded passages. Of its subcategories, ‘Organizational Structures & Change’ (102 units), ‘Resources’ (71 units), ‘Ethics & Work Culture’ (58 units) and ‘Benefits’ (54 units) are the most frequent ones. The least addressed subcategories concern the termination phase of the MSI (8 units) and the aptitude of an MSI as organizational form (9 units). The category ‘Output Effectiveness’ amounts to a total of 10,9 %, ‘Outcome Effectiveness’ to 2,0 % and ‘Impact Effectiveness’ to 3,3% of all coded evaluation questions.
When regarding the categories in relation to the analyzed documents, most (sub)categories are not addressed by at least 1-2 evaluation tools. Exceptions are ‘Organizational Structures & Change’ and ‘Output Effectiveness’, which are comprised by all eleven questionnaires. Standing out is external accountability, which is only included in one document. Internal accountability (5 documents), MSI as organizational form (5 documents), and the termination phase of the MSI (3 documents) are present in less than half of the questionnaires. Inversely, none of the documents comprises all (sub)categories. The number of (sub)categories covered by the analyzed tools ranges from 12 (T5) to 20 (T8) out of the 27 (sub)categories with assigned text passages.

In summary, the analyzed evaluation tools cover a wide range of subjects in their questionnaires, including aspects of legitimacy such as participation, transparency, practices of discourse and decision-making, and accountability mechanisms. They also address various aspects of effectiveness relating to output, outcome, and impact effectiveness, as well as a multitude of factors of institutional effectiveness. In terms of category frequencies, however, some subjects are addressed considerably more often by the evaluation questionnaires than others. The largest number of questions corresponds to the assessment of institutional effectiveness, whereas the fewest questions are directed at evaluating the responsiveness and the outcome and impact effectiveness of the MSI. Although when analyzed together, the evaluation tools comprise a broad spectrum of factors, individual tools do not cover all categories in their assessment.

## 4.2 Further Results of Data Analysis

This chapter provides further results of the data analysis, regarding the overlaps of categorizations and target-specific questions in the investigated evaluation tools.

**Results from Multiple Categorization Feature**

As described in in chapter 3.4.2, I allowed for multiple categorizations during the data analysis. Examining the results of this feature provides further information and insights about the analyzed tools.

Multiple categorizations occur between the two subcategories of 'Transparency' and 'Participation & Inclusiveness', respectively, where either internal and external stakeholders are referred to in the same question, or it is not specified and identifiable from the context whether internal or external stakeholders are addressed (T2, pp. 17, 21; T4; p. 27; T9, pp. 9, 10). Evaluation questions are coded for both, 'Transparency' and 'Participation & Inclusiveness', in cases were a clear distinction between information and participation is not evident from the question (T2, p. 21; T8, pp. 17, 18, 41). There are also overlaps with the category 'Discourse & Deliberative Practice', as these categories are closely linked and some evaluation questions address aspects of more than one category (T1, p. 222; T3, pp. 7, 8; T6, p. 13; T7, p. 109; T8, pp. 16, 24, 59; T9, p. 4; T10, p. 2), such as the provision of information as a basis for members' decision-making (T3, p. 7) or the performance of consultations within the MSI and with external stakeholders (T7, p. 109).

T8 entails several intersections of the categories 'Participation & Inclusiveness' and 'Accountability & Control Mechanisms' when it is assessing which stakeholders form part of certain accountability processes or have access to the control mechanisms (pp. 30, 33, 44, 45). In the same manner, it contains intersections of the categories 'Transparency' and 'Accountability & Control Mechanisms' when detailing the reporting requirements of actors committing to the MSI standards or the transparency of the grievance mechanisms (pp. 33, 45, 48).

Numerous cases of multiple categorizations arise because many categories, despite the definition of a category system (cf. Appendix A), cannot be strictly separated, particularly the subcategories of institutional effectiveness (T1, pp. 221-223; T2, pp. 15, 18, 23; T3, pp. 4, 8, 13; T4, p. 27; T6, pp. 12, 13; T7, p. 110; T8, pp. 24, 54; T9, pp. 3, 9; T10, p. 2; T11, pp. 6, 7). But

also other (sub)categories harbor similar aspects and overlap, for example ‘Participation & Inclusiveness’ and ‘External Support’ (T3, p. 9; T9, p. 9), or ‘Discourse & Deliberative Practice’ and ‘Organizational Structures & Change’ when it comes to the MSI’s structures and processes to ensure a good discursive practice (T1, p. 222; T6, p. 13; T10, p. 2).
However, some passages are coded with multiple categories because one evaluation question encompasses several different aspects, such as “The administrative, communication and decision-making structure of the partnership is clear” (T6, p. 13), “The roles, responsibilities and expectations of partners are clearly defined and understood by all other partners.” (T11, p. 6), “How do the partners plan to work together to achieve these goals?” (T2, p. 15), or “There is a participatory decision-making system that is accountable, responsive and inclusive” (T6, p. 14). Several evaluation questions inquire whether the members have jointly negotiated or agreed upon certain processes, resources, etc. and therefore merge the assessment of two different aspects, the discursive and decision-making process on the one hand and the subject of decision on the other, in one question (T1, p. 222; T2, pp. 16, 18).

Coding overlaps also occur between two or more of the categories of output, outcome, and impact effectiveness as well as the subcategory ‘Vision & Purpose’ (T1, pp. 221, 223; T2, p. 15; T3, pp. 2, 14; T4, p. 27; T5, pp. 19, 39; T8, pp. 12, 14, 24, 56-59; T9, pp. 4, 9; T10, p. 2; T11, pp. 6, 7). This is mainly owed to formulations too unspecific to attribute them to a single category, such as “Does the MSI have a process for reviewing the overall effectiveness of the MSI?” (T8, p. 57), “How effectively does the group meet its aims and objectives?” (T4, p. 27), “Other multiplier effects” (T1, p. 223), or “What is the function of the MSI?” (T8, p. 12). In contrast, some passages in which the term ‘outcome’ is used can be clearly assigned from the context to ‘Output Effectiveness’ according to the category definitions (T1, p. 222; T4, p. 27; T5, p. 19; T8, p. 37; T9, p. 10). Another intersection of coded passages arises between the categories ‘Impact Effectiveness’ and ‘Responsiveness’, which both investigate the MSI’s effect on the target population and affected actors (T3, pp. 3, 13; T4, p. 27; T7, p. 110; T8, p. 13; T9, p. 5).

In conclusion, there are numerous cases of multiple categorizations which mainly occur due to three reasons: (1) Similar and not sharply distinguishable category contents, (2) Mixture of different aspects in one evaluation question, or (3) Use of unspecific or ambiguous terms.

**Specificity of Analyzed Tools towards Target Sector/MSIs**

Most of the analyzed evaluation tools have been developed against the background of a specific sector or tailored to a certain type of MSI (cf. Table 3). In the following paragraph, I outline whether the evaluation questions are specified to a particular sector or type of MSI, or if they are formulated in a universal way.

T1 and T10 are designed for all kinds of MSIs, and their evaluation questions and indicators are therefore generally applicable. T3, T4, T5, T6, T9 and T11 are issued by organizations in the health and public services sector but do not contain any questions specified to this field. T2 is aimed for MSIs in health research (cf. Table 3). It includes several questions specifically directed at research MSIs (pp. 17-18, 21-22, 23), and single references to research in the health sector, such as addressing ethical clearances (p. 16). A large portion of its evaluation questions, however, is of general nature and not targeted to a specific audience. T7 is designed for MSIs in the education sector and its questions are tailored to the target group, especially regarding aspects and indicators of the MSI's relevance and impact (p. 110). T8 is customized for standard-setting MSIs in the field of human rights. Its questionnaire is focused on human rights, and large parts of it concern standards setting and control mechanisms. Yet, it leaves open the industry it is applied to.

In summary, eight of the analyzed eleven tools do not include evaluation questions customized to target audiences. The other three evaluation questionnaires are specified for particular user groups to varying extents.

# 5 Discussion

In this chapter, I discuss the results from the empirical research presented in the previous chapter in relation to the theoretical background provided in chapter 2 and my research questions. Subchapter 5.1 answers the research questions posed in this master thesis based on the discussion and interpretation of the empirical results, and briefly describes the limitations of this research. Subchapter 5.2 examines the implications of the research findings for the field of human resources and organizational development. The final subchapter explores the outlook and recommendations for future research.

## 5.1 Discussion and Interpretation of the Empirical Results

This master thesis aims to add to the knowledge and improvement of the evaluation of MSIs, especially in the climate sector. It makes the following contribution to the literature: By means of systematically analyzing current MSI evaluation tools, it shows that the tools largely reflect the deficits of MSIs, and that although they cover some evaluation aspects appropriately, they are not suitable to comprehensively evaluate MSIs in general and climate-related MSIs in particular.

In this chapter, I discuss my research results against the theoretical background of the current state of scientific knowledge (see chapter 2) to answer the research questions of this master thesis. The chapter is subdivided in four parts, addressing the two sub-questions and the main research question, as well as briefly outlining the limitations of this research.

### 5.1.1 Evaluation Aspects Covered by Current Tools Compared to Scientific Literature

The first sub-question of my research investigates which aspects are already covered by the currently available evaluation tools and how these compare to scientific literature.

The analysis shows that currently existing evaluation tools largely parallel the deficits of MSIs and their evaluation criticized by researchers (cf. e.g. Andonova & Levy, 2003; Bäckstrand, 2012; Biermann et al., 2007; Pattberg & Widerberg, 2016). At first glance, it appears that current evaluation tools cover the indicators of legitimacy and effectiveness stated in literature (cf. Bäckstrand, 2012; Bernstein, 2011; Dingwerth, 2007; Jastram & Klingenberg, 2018; Wolf, 2010). However, a closer examination of the exact formulations and frequencies of the evaluation questions reveals shortcomings and limitations.

Looking into the aspects related to the three sources of democratic legitimacy according to Dingwerth (2007), the results for participation and inclusiveness show that nine out of eleven analyzed tools comprise this category in their questionnaire (cf. Table 5). In reference to the three main dimensions to assess participation stated by Bäckstrand (2012), only one tool addresses the geographical dimension of internal and external participation (cf. T8). While a single tool specifically assesses the participation of non-governmental groups as MSI members (cf. T8), several tools include the dimension of non-governmental participation when evaluating the external participation of stakeholders (cf. T2; T4; T6; T8; T9). The participation of marginalized groups is not addressed by any of the investigated evaluation tools. In terms of specification, several tools refer to the inclusion of the 'relevant' or 'right' actors without defining which actors these are, subjecting this judgement to the bias of the persons completing the questionnaire (cf. e.g. T4; T6). Although many evaluation tools determine which stakeholders are part of the MSI and whether they contribute the necessary skills to the initiative, they do not specify which actors should be and are not participating as members or external stakeholders. Therefore, the analyzed evaluation tools appear to be inadequate for assessing whether the MSI is reducing or perpetuating the existing participation gap (cf. Bäckstrand, 2012; Biermann et al., 2007), which is one of the largest doubts about MSIs (e.g. Andonova & Levy, 2003; Biermann et al., 2007).

With respect to democratic control as second source of democratic legitimacy (Dingwerth, 2007), some aspects of transparency are comprised by ten out of eleven analyzed questionnaires (cf. Table 5). Most of the evaluation questions determine which information is shared with which stakeholder group(s) via which channels. This corresponds to the dimension of providing and distributing information on part of the MSI, while the dimension of resources to access and process them on part of the audience (cf. Bäckstrand & Kylsäter, 2014; Dingwerth, 2007) is left out by most evaluation tools. A single tool addresses the language(s) of the target group(s) and the available information, as well as the geographic location(s) of information program offers (cf. T8). Some questionnaires assess whether the MSI's purpose and planned actions are clear to the public, which also implies the resources to access the provided information, but they include no further specifications (cf. T3; T4; T10). Concerning accountability, five out of eleven evaluation tools address the issue at all, and only one of them investigates accountability towards external stakeholders (cf. Table 5). Even when addressing accountability, some evaluation tools do it in a very generalized way, such as inquiring whether the MSI structures are designed to support accountability, without further investigating factors or mechanisms of

accountability (cf. T5; T6; T11). While current evaluation tools partly encompass transparency aspects, their coverage of accountability mechanisms is, with one exception (cf. T8), slight or nonexistent, which coincides with the criticism of lack of accountability mechanisms and of accountability in MSI practice (e.g. Bäckstrand & Kuyper, 2017; Hale & Mauzerall, 2004). Responsiveness, although the ultimate target of functioning transparency and accountability mechanisms (cf. Dingwerth, 2007), is little addressed and investigated according to my literature research (cf. chapter 2), which is also reflected in the analyzed evaluation questionnaires. Even though responsiveness indicates that an MSI is indeed acting in the interest of the affected actors and beneficial to the collective action problem at hand (cf. Dingwerth, 2007), four of the eleven analyzed tools do not address this issue in their evaluation questions (cf. Table 5). However, some of the tools that examine responsiveness do so in detailed and specified questions (cf. T7; T8) and continue to question the MSI's responsiveness during the implementation phase (cf. T4, T10).

Regarding the MSI's discursive practice, the third source of democratic legitimacy (Dingwerth, 2007), the analyzed evaluation tools address various factors mentioned in literature (cf. Bäckstrand, 2012; Dingwerth, 2007; Mena & Palazzo, 2012), such as equal participation in the decision-making process, power relations, and the discussion of opposing views (cf. T1; T3; T6; T7; T8). Although two of the analyzed questionnaires do not assess the discursive practice of the MSI (cf. Table 5), overall it is adequately covered by current tools, and the results of my research do not resemble Bäckstrand's (2012) criticism that the deliberation process within MSIs is not examined.

My hypothesis for this discrepancy is that researchers might not have easy access to investigate MSI-internal discursive and deliberation processes, while the analyzed evaluation tools are designed for the self-assessment of MSIs, where these barriers do not exist.

Looking at the different dimensions of effectiveness of MSIs (cf. chapter 2.6.3), the results for output effectiveness show that evaluation tools mainly explore whether the MSI has jointly agreed and clearly defined goals (cf. T1; T2; T3; T4; T5; T6; T10; T11). However, most tools do not delve into the metric of evaluation (cf. Miles et al., 2002); only two of them investigate the indicators and data necessary for measuring progress towards the goals (cf. T4; T9). The fact that, of the three dimensions, the output dimension is covered most extensively by the analyzed tools (see Table 5) also reflects that it is easier to determine than outcome and impact effectiveness (cf. Wolf, 2010).

Outcome effectiveness is hardly addressed by the analyzed evaluation questionnaires. Even though I categorized evaluation questions regarding capacity and awareness building (cf. T2; T8) as 'outcome effectiveness', capacity and awareness building might, but do not necessarily entail behavioral changes, and therefore do not even correspond to this category if strictly following definition (cf. Jastram & Klingenberg, 2018; Wolf, 2010). Most other questions are assigned to this category due to vague terminology, e.g. 'objectives' or 'overall effectiveness', which makes it difficult to unambiguously attribute them to one category – a challenge that applies to all dimensions of effectiveness.

Concerning the impact effectiveness of the MSI, several tools ask whether the initiative makes a contribution to the community (cf. T3; T6; T11). However, they mostly do not question who the aspired beneficiaries and the affected actors are – only two tools do (cf. T8; T9) – and equate 'contribution to the community' with positive impact. Just two questionnaires specify what the aspired contribution or impact is and what its indicators are (cf. T7; T8). A single tool addresses possible adverse effects the MSI might cause (cf. T8). In reference to Wolf's (2010) approach to the evaluation of impact effectiveness, the evaluation questions address the perspective of 'corporate satisfaction' but do not cover the perspectives of 'stakeholder satisfaction' and researchers' satisfaction', both of which would need to be included in a comprehensive evaluation of impact effectiveness.

Institutional effectiveness is mentioned in literature as either a part of output effectiveness (cf. Wolf, 2010) or an additional dimension of the effectiveness of organizations (cf. Bäckstrand, 2006; 2012), and according to my research of scientific literature (cf. chapter 2) it is not attributed much importance in the context of evaluation requirements for MSIs. In my empirical research, I had initially entered institutional effectiveness into the analysis as one of nine deductively derived categories (cf. Table 4). The number and scope of corresponding evaluation questions during the preliminary coding phase (cf. chapter 3.4.2), however, lead to the inductive formation of twelve subcategories, and ultimately, the results showed that 46 % of all coded text passages refer to aspects of institutional effectiveness (cf. Table 5). The analyzed questionnaires encompass a wide range of evaluation aspects in this category, as presented in chapter 4.1, and although not every aspect is comprised by every tool, they generally cover the evaluation of institutional effectiveness well and in-depth. Rather, the results suggest that this dimension is accorded disproportionate importance in the evaluations. While main concerns and criticism of researchers are the lack of (evidence for the) legitimacy and effectiveness of MSIs in addressing collection action problems (cf. e.g. Bäckstrand, 2012; Biermann et al., 2007; Mena & Palazzo, 2012; Pattberg & Widerberg, 2016), the analyzed evaluation tools direct almost half

of the questions to matters of internal structures and designs, and appear to give them a higher priority than whether the MSI is fulfilling its function of compensating deficits in global governance (cf. e.g. Bäckstrand, 2012; Biermann et al., 2007).
My hypothesis for this divergence between the academic literature and existing evaluation tools is that the researchers' perspective focusses more on the supposed functions of MSIs and the evaluation of their attainment, while evaluation tools mainly reflect organizations' interests and practical considerations that prioritize their own working conditions and benefits, as well as aspects that can be assessed with moderate effort.

In summary, the results indicate that the currently existing evaluation tools for the most part reflect the deficits of MSIs criticized in the literature. While a range of evaluation questions address sources of MSI legitimacy, they often examine only the MSI's current practices and do not compare them to benchmarks or the recipients' side. Examples include identifying the MSI members without comparing them to dimensions for assessing participation (cf. Bäckstrand, 2012), or exploring the provision of information but not the resources of the target audience to access and process that information. Particularly striking is the lack of evaluation of accountability mechanisms, especially towards external actors, which mirrors the strong criticism that accountability mechanisms for MSIs are almost completely absent (cf. e.g. Bäckstrand & Kuyper, 2017). Regarding the effectiveness of MSI operation, currently available tools extensively cover the evaluation of institutional effectiveness. Output effectiveness is less explored in the analyzed questionnaires, and specifics regarding the indicators and measurement are mostly missing. Outcome and impact effectiveness are addressed even less, possibly corresponding to the fact that they are more complex and elaborate to determine (cf. Wolf, 2010). The prioritization of the different dimensions of effectiveness in the analyzed tools is opposed to their collective relevance (cf. Jastram & Klingenberg, 2018; Wolf, 2010).

### 5.1.2 Limitations of the Applicability of Current Tools

The second sub-question of my research addresses limitations regarding the applicability of current evaluation tools, in general and for climate-related MSIs in particular. In this context, I discuss the results from the multiple categorization feature and the specificity of the tools as well as additional relevant insights from the previous subchapter.

The results show a lack of unequivocal terminology. On the one hand, in many cases tools do not specify the terminology enough. For example, when using terms like ‘aims and objectives’, ‘overall effectiveness’ or ‘function of the MSI’ (cf. T4; T8), it is not discernible whether they refer to the mission, output or impact goals of the MSI. On the other hand, an ambiguous use of terminology also results from the different possibilities of classifying MSI effects, as discussed in chapter 2.6.4, for example, when the term ‘outcome’ in a questionnaire (cf. T1; T4; T5; T8; T9) would be classified as ‘output’ according to Jastram and Klingenberg (2018) and Wolf (2010).

A further reason for unspecific evaluation questions is that different evaluation aspects are combined in one question, which does not allow for a clear assessment of their components, e.g., when asking about the result and the process of a decision or about different characteristics at the same time (cf. T1; T2; T6; T11).

Another drawback is that the analyzed tools in many cases lack specific indicators of the evaluation aspects, leaving it to the individuals completing the questionnaire to judge whether a certain aspect has been met or accomplished. This shortcoming extends from indicators for an inclusive participation or for determining the accountability of a process to indicators for the achievement of output, outcome, and impact goals (cf. T2; T3; T4; T6; T9; T11). Particularly with regard to the impact dimension, the absence of indicators is also related to the problem raised by researchers of who defines the relevant criteria and who is in the position to assess them (cf. Wolf, 2010). For the issue of participation on the other side, assessment dimensions have already been suggested in literature (cf. Bäckstrand, 2012). The lack of indicators and the subsequent assessment according to subjective perception also shows parallels to the criticism that reporting is mainly based on self-description and ultimately relates back to a lack of (external) accountability (cf. Bäckstrand & Kuyper, 2017).

Regarding the area of climate governance, none of the analyzed evaluation tools (nor of the initially gathered tools, cf. chapter 3.4.1) was designed for climate-related MSIs. However, eight of the eleven tools do not include specified evaluation questions for a particular sector and have the advantage of being universally applicable (cf. T3; T4; T5; T6; T9; T10; T11). The tools that are tailored to certain target users, on the other hand, include specific indicators, particularly regarding the impact dimension of the MSI (cf. T7; T8). These sector-specific indicators would need to be converted into assessment criteria for climate protection efforts, which revisits the problem of defining the relevant criteria mentioned in the previous paragraph (cf.

Wolf, 2010), which is an ongoing debate for climate governance (cf. Ven et al., 2017; Widerberg & Pattberg, 2015).
In addition, most analyzed tools do not include a determination of the MSI's target population and affected actors (cf. T1; T2; T3; T4; T5; T6; T7; T10; T11), which would be very relevant for climate-related MSIs. In global climate governance, MSIs face an even bigger challenge identifying and including the beneficiaries and affected actors than in other areas, as their range is especially extensive and complex (cf. Andonova et al., 2009).

In conclusion, most current evaluation tools can be applied to climate-related MSIs without adaptions, although the tools in general have some deficits regarding the precision of terminology and questions as well as the provision of concrete indicators. To meet the needs of MSIs in global climate governance more appropriately, some additional questions and indicators would be necessary.

### 5.1.3 Suitability of Current Evaluation Tools for MSIs in Global Climate Governance

The overall research question of my work examines the suitability of current evaluation tools for the evaluation of MSIs in global climate governance.
The results of my empirical research show that while my data collection did not reveal any evaluation tool designed for climate-related MSIs, most currently available tools do not include questions specified for a particular sector and are applicable to MSIs in climate governance without adaptions. The analyzed evaluation questionnaires, however, mirror the deficits of MSIs and their evaluation stated in scientific literature (cf. Bäckstrand, 2012; Bernstein, 2011; Dingwerth, 2007; Jastram & Klingenberg, 2018; Wolf, 2010). For example, the criticized absence of accountability mechanisms (cf. Bäckstrand & Kuyper, 2017; Hale & Mauzerall, 2004) or the reinforcement of the existing power imbalances and participation gap (cf. Andonova & Levy, 2003; Biermann et al., 2007) are reflected in the evaluation questionnaires through the absence of corresponding assessment questions. In addition, climate-related MSIs face particular challenges due to the complexity of the issue (cf. Andonova et al., 2009) and the unresolved question of adequate impact criteria (cf. Ven et al., 2017; Widerberg & Pattberg, 2015). For example, climate governance efforts are confronted with a particular wide range of affected actors and with coordination requirements across borders, sectors, hierarchies etc. (cf. Andonova et al., 2009), all of which are not sufficiently represented in the available questionnaires.

In conclusion, the results show that current evaluation tools mirror the deficits of MSIs and their evaluation criticized by academics, and are not sufficiently equipped to evaluate the contribution to the participation gap, accountability mechanisms and the impact of MSIs. Also, they do not include certain aspects specific to MSIs in climate governance. However, the currently available tools might serve as a good starting point for evaluation and are suitable to cover some topics such as transparency, discourse and decision-making practice, as well as institutional effectiveness.

### 5.1.4 Limitations of the Research

One limitation of this research is that the data sample could not be determined along fixed search parameters and is therefore not objectively replicable, as mentioned in chapter 3.4.1. Another severe drawback is that the analyzed evaluation tools are all published by organizations or authors based in Europe, North America, and Australia, and though some of them operate or cooperate globally, this replicates the very participation gap MSIs are supposed to reduce (cf. Bäckstrand, 2012; Biermann et al., 2007).
Also, this analysis of the suitability and applicability of current evaluation tools does not provide any information as to if and how MSIs employ them in practice and whether they draw consequences from the results. Moreover, the present work does not delve into the background, interests, and scientific foundation of the development of each of the analyzed tools.

## 5.2 Implications for Human Resource and Organizational Development

Various aspects that are relevant from an HR perspective are only covered to a limited extent by the currently existing MSI evaluation tools. Some evaluation tools assess whether the persons involved in the MSI possess the necessary skills and expertise for its collaborative action (cf. T3; T6; T9). How additionally needed skills or knowledge can be acquired is not addressed by any of the analyzed questionnaires, nor are training and development programs for employees in the context of the MSI. Processes to ensure learning among members during the MSI operation are only included by one evaluation tool (cf. T6), and none of the questionnaires cover the transfer of learnings and knowledge into the member organizations during the implementation and after the finalization of the MSI. Also, the (re-)deployment of staff after the termination of the MSI is addressed by a single evaluation tool (cf. T2). For a sustainable HR management and development, these topics need to be considered in advance before they come into effect. This means that aspects relevant for HR management should be included in evaluation

questionnaires, but primarily should also be taken into consideration in the MSI practice, and HR experts should be involved before even deciding to engage in an MSI.
As the MSI forms an organization on its own, with its own vision, values, roles, and structures, it also has its own HR issues. Evaluation tools address various structures and processes of the MSI, such as governance, leadership, or communication arrangements (cf. T1; T2; T3; T6; T8; T10; T11), but do not address how the HR issues of the MSI are managed. Depending on its size and duration, the MSI might need its own HR management, which should be considered when establishing and structuring the MSI and should also be part of the MSI evaluation.
The MSI as an organization of its own is also relevant on a level of organizational development. Aspects like (building) the organizational identity of the MSI, the working culture, and the management of changes are addressed in the evaluation tools. However, the issue of how is being dealt in day-to-day operation with the two organizations – the member organization and the MSI – and their identities, cultures, and structures and possible differences between them, is not addressed. One evaluation questionnaire includes the compatibility of core values and cultures of the member organizations as an indicator (cf. T1), but they will never be congruent. This discrepancy should be actively addressed and managed, particularly in the member organizations and if the engagement in the MSI is of major extent, and it should also be considered in the evaluation.
Outside of a MSI involvement, from an organizational development point, the respective parts of MSI evaluation tools could also be applied and valuable for evaluating the own organization.

In summary, the analysis of current MSI evaluation tools shows that topics of HR management are only marginally addressed. On the one hand, this should be considered in the composition of evaluation questionnaires. More importantly, assuming that the evaluation tools mirror the importance attributed to certain areas, this means that the relevance of HR management in the context of MSIs is underrated. HR experts and issues ought to be more involved when of engaging in collective action initiatives.

## 5.3 Outlook and Recommendations

This master thesis shows that current evaluation tools are not suitable to comprehensively evaluate the legitimacy and effectiveness of MSIs, and subsequently the fulfillment of their main functions. The development of an adequate evaluation tool for MSIs in general and for climate-related MSIs in particular is therefore a subject for further research. This research would need

to incorporate scientific knowledge and researchers' criticism into the development process, to avoid that the shortcomings of MSIs are reflected in the evaluation questionnaire. It should also be investigated in a field phase, however, how well the developed tool works and is manageable for users in practice. In order to design an evaluation tool for MSIs in climate governance, additional research and discourse are necessary regarding the specific indicators for the impact assessment of climate protection efforts, since researchers disagree on the subject and many open questions and debates remain (cf. e.g. Ven et al., 2017; Widerberg & Pattberg, 2015). In reference to the limitations of this work, studies should also be conducted on the evaluation practice of MSIs and whether performed evaluations lead to changes in the MSIs and their activities.

This master thesis reveals a divergence between the prioritization of institutional effectiveness in the analyzed questionnaires and the prioritization of impact effectiveness in scientific opinion and politics (cf. Jastram & Klingenberg, 2018; Wolf, 2010). This discrepancy and the interests behind each prioritization are also an area for future research.

Another objective for further investigation is to gain a deeper understanding of the behavior-related aspects of MSI evaluation, which are often more difficult to access and determine. This concerns particularly the deliberative and discursive practice of MSIs, which is strongly represented in the analyzed evaluation tools, but little studied (cf. Bäckstrand, 2012), and the outcome effectiveness of MSIs, which is scarcely represented both in evaluation tools and research (cf. e.g. Jastram & Klingenberg, 2018).

In general, the evaluation of MSIs, particularly in the climate sector, is a field of current interest, and the state of knowledge and research will be continuously evolving and opening up new requirements and possibilities for further research.

# 6 Summary

Multi-stakeholder initiatives (MSIs) are collaborative, cross-sectoral efforts between various actors that target current social or environmental challenges and are considered to play an important role in achieving the Sustainable Development Goals (cf. e.g. Pattberg & Widerberg, 2016; United Nations, 2015b). MSIs, being based on voluntary compliance and non-governmental structures, require legitimacy in addition to effectiveness in fulfilling their functions (cf. e.g. Bäckstrand, 2012; Gregorio et al., 2020; Mena & Palazzo, 2012; Pattberg & Widerberg, 2016). Critics increasingly question this legitimacy and effectiveness, and there is little evidence for the successful performance of MSIs (cf. Bäckstrand, 2012; Biermann et al., 2007; MSI Integrity, 2020a; Pattberg & Widerberg, 2016). It is essential to evaluate MSIs to ensure their legitimacy and effectiveness as well as their contribution to pressing global problems (cf. OECD, 2008; Stadtler, 2016). But to date, the knowledge and research on the evaluation of MSIs are limited and few studies have systematically investigated the evaluation of collective action efforts (e.g. Biermann et al., 2007; Stadtler, 2016; Ven et al., 2017). Even though several tools and frameworks have been developed to facilitate the evaluation of MSIs, there is barely any research analyzing these evaluation tools and their adequacy for this purpose, a problem which my work aims to address.

By means of systematically analyzing currently available MSI evaluation tools, this master thesis provides new findings about the suitability and applicability of current tools for MSIs in the climate sector. I direct my research at MSIs in global climate governance because of their particular multi-stakeholder character and the pressing urgency of the problem (cf. Andonova et al., 2009; Jagers & Stripple, 2003).

To conduct my empirical research, I employed a qualitative comparative study. I used document analysis for the data collection of this thesis. For the data analysis, I applied a qualitative content analysis according to Mayring with a deductive category assignment, which permitted me to establish an initial analytical frame for the strongly variable entry data and to flexibly integrate newly acquired knowledge in the course of the analytical process (cf. Mayring, 2014).

The results show that the analyzed evaluation tools cover a wide range of subjects, including aspects of legitimacy such as participation, decision-making practices, and accountability mechanisms, along with aspects relating to output, outcome, and impact effectiveness, as well as a multitude of factors of institutional effectiveness. By far the largest number of evaluation questions corresponds to the assessment of institutional effectiveness. Ambiguity in the category assignment occurred in numerous cases, either because of similar category contents, or

due to different evaluation aspects included in one question, or because of unspecific or ambiguous terminology. With regards to the customization of the analyzed tools to a target audience, eight questionnaires show no specificity while three are customized to a particular target group to varying extents.

In conclusion, my research shows that most current evaluation tools could be applied to climate-related MSIs without adaption. The analyzed tools, however, mirror the deficits of MSIs as stated in scientific literature (cf. Bäckstrand, 2012; Bernstein, 2011; Dingwerth, 2007; Jastram & Klingenberg, 2018; Wolf, 2010). For example, they are insufficiently equipped to assess the impact effectiveness of an MSI or whether it contributes to closing the participation gap (cf. Andonova & Levy, 2003; Biermann et al., 2007). Particularly striking is the lack of assessment of accountability mechanisms, especially towards external actors. Furthermore, the evaluation questionnaires show some deficits regarding their precision of questions and terminology, and the provision of concrete indicators. Also, the challenges and requirements of MSIs in global climate governance are not sufficiently addressed by available questionnaires. Some issues, however, are quite well-covered, such as transparency or discourse and decision-making practices. Institutional effectiveness is particularly extensively addressed and seems to constitute the major priority of organizations participating in MSIs. Even though the analyzed evaluation tools have various deficits, they might serve as a starting point for the evaluation of MSIs in the climate sector and are suitable to cover some topics.

# 7 References


Africa Liaison Program Initiative. (n. d.). *Ensuring Successful Partnerships: A Toolkit*. Washington, DC. https://portailqualite.acodev.be/fr/system/files/node/197/ensuring_successful_partnerships.pdf

Afsana, K., Habte, D., Hatfield, J., Murphy, J., & Neufeld, V. (2009). *Partnership Assessment Toolkit.* Canadian Coalition for Global Health Research. https://www.elrha.org/wp-content/uploads/2014/08/PAT_Interactive_e-1.pdf

Andonova, L. B., Betsill, M. M., & Bulkeley, H. (2009). Transnational Climate Governance. *Global Environmental Politics*, *9*(2), 52–73.

Andonova, L. B., & Levy, M. A. (2003). Franchising Global Governance: Making Sense of the Johannesburg Type II Partnerships. In O. S. Stokke & Ø. B. Thommessen (Eds.), *International environmental governance: Volume 18. Yearbook of international cooperation on environment and development 2003-04* (pp. 19–31). Earthscan Publications.

Arenas, D., Albareda, L., & Goodman, J. (2020). Contestation in Multi-Stakeholder Initiatives: Enhancing the Democratic Quality of Transnational Governance. *Business Ethics Quarterly*, *30*(2), 169–199. https://doi.org/10.1017/beq.2019.29

Bäckstrand, K. (2006). Multi-stakeholder partnerships for sustainable development: rethinking legitimacy, accountability and effectiveness. *European Environment*, *16*(5), 290–306. https://doi.org/10.1002/eet.425

Bäckstrand, K. (2012). Are partnerships for sustainable development democratic and legitimate? In P. H. Pattberg (Ed.), *Public-private partnerships for sustainable development: Emergence, influence and legitimacy* (pp. 165–182). Elgar.

Bäckstrand, K., & Kuyper, J. W. (2017). The democratic legitimacy of orchestration: the UNFCCC, non-state actors, and transnational climate governance. *Environmental Politics*, *26*(4), 764–788. https://doi.org/10.1080/09644016.2017.1323579

Bäckstrand, K., & Kylsäter, M. (2014). Old Wine in New Bottles? The Legitimation and Delegitimation of UN Public–Private Partnerships for Sustainable Development from the Johannesburg Summit to the Rio+20 Summit. *Globalizations*, *11*(3), 331–347. https://doi.org/10.1080/14747731.2014.892398

Bakker, F. G. de, Rasche, A., & Ponte, S. (2019). Multi-Stakeholder Initiatives on Sustainability: A Cross-Disciplinary Review and Research Agenda for Business Ethics. *Business Ethics Quarterly*, *29*(03), 343–383. https://doi.org/10.1017/beq.2019.10

Barkemeyer, R., Preuss, L., & Lee, L. (2015). On the effectiveness of private transnational governance regimes—Evaluating corporate sustainability reporting according to the Global Reporting Initiative. *Journal of World Business*, *50*(2), 312–325. https://doi.org/10.1016/j.jwb.2014.10.008

Bernstein, S. F. (2011). Legitimacy in intergovernmental and non-state global governance. *Review of International Political Economy*, *18*(1), 17–51. https://doi.org/10.1080/09692290903173087

Biekart, K., & Fowler, A. (2018). Ownership dynamics in local multi-stakeholder initiatives. *Third World Quarterly*, *39*(9), 1692–1710. https://doi.org/10.1080/01436597.2018.1450139

Biermann, F., Chan, M., Mert, A., & Pattberg, P. H. (2007). Multi-stakeholder Partnerships for Sustainable Development: Does the Promise Hold? In A. P. J. Mol, F. Biermann, & P. Glasbergen (Eds.), *Partnerships, governance and sustainable development: Reflections on theory and practice.* Edward Elgar.

Bradford, A. (2007). Regime Theory. In *Max Planck Encyclopedia of Public International Law*. Oxford University Press. https://scholarship.law.columbia.edu/cgi/viewcontent.cgi?article=2971&context=faculty_scholarship

Brinkerhoff, D. W., & Brinkerhoff, J. M. (2011). Public-private partnerships: Perspectives on purposes, publicness, and good governance. *Public Administration and Development*, *31*(1), 2–14. https://doi.org/10.1002/pad.584

Brinkerhoff, J. M. (2002). Assessing and improving partnership relationships and outcomes: a proposed framework. *Evaluation and Program Planning*, *25*(3), 215–231. https://doi.org/10.1016/S0149-7189(02)00017-4

Buchanan, A., & Keohane, R. O. (2006). The Legitimacy of Global Governance Institutions. *Ethics & International Affairs*, *20*(4), 405–437. https://doi.org/10.1111/j.1747-7093.2006.00043.x

Bulkeley, H., Andonova, L. B., Bäckstrand, K., Betsill, M. M., Compagnon, D., Duffy, R., Kolk, A., Hoffmann, M. J., Levy, D., Newell, P., Milledge, T., Paterson, M., Pattberg, P. H., & VanDeveer, S. D. (2012). Governing Climate Change Transnationally: Assessing the Evidence from a Database of Sixty Initiatives. *Environment and Planning C: Government and Policy*, *30*(4), 591–612. https://doi.org/10.1068/c11126

Bulkeley, H., & Schroeder, H. (2012). Beyond state/non-state divides: Global cities and the governing of climate change. *European Journal of International Relations*, *18*(4), 743–766. https://doi.org/10.1177/1354066111413308

Cabaj, M. (2014). Evaluating Collective Impact: Five Simple Rules. *The Philanthropist*, *26*(1), 109–124. https://www.tamarackcommunity.ca/hubfs/Resources/Publications/Evaluating%20Collective%20Impact%205%20Simple%20Rules.pdf

Cambridge University Press. (2014). *Cambridge Dictionary*. https://dictionary.cambridge.org/de/worterbuch/englisch

Center for the Advancement of Collaborative Strategies in Health. (2002). *Partnership Self-Assessment Tool*. https://atrium.lib.uoguelph.ca/xmlui/bitstream/handle/10214/3129/Partnership_Self-Assessment_Tool-Questionnaire_complete.pdf?sequence=1&isAllowed=y

Chianca, T. (2008). The OECD/DAC Criteria for International Development Evaluations: An Assessment and Ideas for Improvement. *Journal of MultiDisciplinary Evaluation*, *5*(9), 41–51.

Clarke, A., & MacDonald, A. (2019). Outcomes to Partners in Multi-Stakeholder Cross-Sector Partnerships: A Resource-Based View. *Business & Society*, *58*(2), 298–332. https://doi.org/10.1177/0007650316660534

Compassion Capital Fund National Resource Center. (2010). *Partnerships: Frameworks for Working Together* (Strengthening Nonprofits: A Capacity Builder's Resource Library). https://cwdb.ca.gov/wp-content/uploads/sites/43/2020/12/A-Capacity-Builders-Resource-Library-Partnerships-Frameworks-for-Working-Together_ACCESSIBLE.pdf

Dingwerth, K. (2007). *The new transnationalism: Transnational governance and democratic legitimacy*. *Transformations of the state*. Palgrave Macmillan.

Domínguez, R., & Velázquez Flores, R. (2012). Global Governance. In B. Meiches & R. Hopkins (Eds.), *Oxford Research Encyclopedia of International Studies.* Oxford University Press. https://doi.org/10.1093/acrefore/9780190846626.013.508

Easton, D. (1965). *A systems analysis of political life*. Wiley.

ESTHER Alliance for Global Health Partnerships. (n. d.). *EFFECt Tool: Self Assessment Version: EFFective in Embedding Change tool*. https://www.esther.ie/wp-content/uploads/2021/09/Questionnaire_effect_tool_ESTHER-Alliance.pdf

Eweje, G., Sajjad, A., Nath, S. D., & Kobayashi, K. (2021). Multi-stakeholder partnerships: a catalyst to achieve sustainable development goals. *Marketing Intelligence & Planning*, *39*(2), 186–212. https://doi.org/10.1108/MIP-04-2020-0135

Flick, U. (2004). Design and Process in Qualitative Research. In U. Flick, E. von Kardorff, & I. Steinke (Eds.), *A companion to qualitative research.* SAGE Publications.

Flick, U., Kardorff, E. von, & Steinke, I. (2004). What is Qualitative Research? An Introduction to the Field. In U. Flick, E. von Kardorff, & I. Steinke (Eds.), *A companion to qualitative research.* SAGE Publications.

Fransen, L. (2012). Multi-stakeholder governance and voluntary programme interactions: legitimation politics in the institutional design of Corporate Social Responsibility. *Socio-Economic Review*, *10*(1), 163–192. https://doi.org/10.1093/ser/mwr029

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Aldine.

Global Development Incubator. (2015). *More Than The Sum Of Its Parts: Making Multi-Stakeholder Initiatives Work.* Global Development Incubator. https://globaldevincubator.org/wp-content/uploads/2016/02/Making-MSIs-Work.pdf

Global Reporting Initiative. (2022). https://www.globalreporting.org

Göbel, T. (2010). *Decent Work and Transnational Governance: Multi-stakeholder initiatives' impact on labour rights in global supply chains* (1st ed.). Nomos Verlagsgesellschaft mbH & Co. KG.

Grant, R. W., & Keohane, R. O. (2005). Accountability and Abuses of Power in World Politics. *American Political Science Review*, *99*(1), 29–43. https://doi.org/10.1017/S0003055405051476

Gregorio, M. Di, Massarella, K., Schroeder, H., Brockhaus, M., & Pham, T. T. (2020). Building authority and legitimacy in transnational climate change governance: Evidence from the Governors' Climate and Forests Task Force. *Global Environmental Change*, *64*. https://doi.org/10.1016/j.gloenvcha.2020.102126

Guerra-López, I. J. (2008). *Performance Evaluation: Proven Approaches for Improving Program and Organizational Performance*. John Wiley & Sons, Incorporated.

Gullickson, A. M. (2020). The whole elephant: Defining evaluation. *Evaluation and Program Planning*, *79*, 1–9. https://doi.org/10.1016/j.evalprogplan.2020.101787

Hale, T. N., & Mauzerall, D. (2004). Thinking Globally and Acting Locally: Can the Johannesburg Partnerships Coordinate Action on Sustainable Development? *The Journal of Environment & Development*, *13*(3), 220–239. https://doi.org/10.1177/1070496504268699

Hale, T. N., & Roger, C. B. (2014). Orchestration and transnational climate governance. *The Review of International Organizations*, *9*(1), 59–82. https://doi.org/10.1007/s11558-013-9174-0

Hardy, B., Hudson, B., & Waddington, E. (2003). *Assessing Strategic Partnership: The Partnership Assessment Tool*. London. Office of the Deputy Prime Minister of the United Kingdom; Nuffield Institute of Health. https://www.researchgate.net/publication/265039661_Assessing_Strategic_Partnership_The_Partnership_Assessment_Tool

Hens, L., & Nath, B. (2003). The Johannesburg Conference. *Environment, Development and Sustainability*, *5*(1/2), 7–39. https://doi.org/10.1023/A:1025303511864

Hoffmann, N. (2012). Dokumentenanalyse. In B. Schäffer & O. Dörner (Eds.), *Handbuch Qualitative Erwachsenen- und Weiterbildungsforschung* (pp. 395–406). Barbara Budrich.

Jagers, S. C., & Stripple, J. (2003). Climate governance beyond the state. *Global Governance*, *9*(3), 385–399.

Jastram, S. M., & Klingenberg, J. (2018). Assessing the Outcome Effectiveness of Multi-Stakeholder Initiatives in the Field of Corporate Social Responsibility – The Example of the United Nations Global Compact. *Journal of Cleaner Production*, *189*, 775–784. https://doi.org/10.1016/j.jclepro.2018.04.005

Keohane, R. O. (2011). Global governance and legitimacy. *Review of International Political Economy*, *18*(1), 99–109. https://doi.org/10.1080/09692290.2011.545222

Liu, H. J., Love, P. E. D., Smith, J., Sing, M. C. P., & Matthews, J. (2018). Evaluation of public–private partnerships: A life-cycle Performance Prism for ensuring value for money. *Environment and Planning C: Politics and Space*, *36*(6), 1133–1153. https://doi.org/10.1177/2399654417750879

Louis, W. R. (2009). Collective Action-and Then What? *Journal of Social Issues*, *65*(4), 727–748. https://doi.org/10.1111/j.1540-4560.2009.01623.x

Mahmood, S., Morreale, S., & Barry, M. M. (2015). *Developing a Checklist for Intersectoral Partnerships for Health Promotion.* National University of Ireland Galway, World Health Organization Collaborating Centre for Health Promotion Research. https://www.researchgate.net/publication/280573639_Developing_a_Checklist_for_Intersectoral_Partnerships_for_Health_Promotion

Marriott, N., & Goyder, H. (2009). *Manual for monitoring and evaluating education partnerships*. International Institute for Educational Planning. https://unesdoc.unesco.org/ark:/48223/pf0000185117

Mayring, P. (2014). Qualitative content analysis: theoretical foundation, basic procedures and software solution. *Social Science Open Access Repository (SSOAR)*, 1–143. https://www.ssoar.info/ssoar/handle/document/39517

Mayring, P. (2015). *Qualitative Inhaltsanalyse: Grundlagen und Techniken* (12th, revised ed.). Beltz.

Mayring, P. (2019). Qualitative Content Analysis: Demarcation, Varieties, Developments. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, *20*(3).

Mayring, P. (2020). Qualitative Inhaltsanalyse. In G. Mey & K. Mruck (Eds.), *Handbuch Qualitative Forschung in der Psychologie* (pp. 495–511). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-26887-9_52

Mayring, P., & Fenzl, T. (2014). Qualitative Inhaltsanalyse. In N. Baur & J. Blasius (Eds.), *Handbuch Methoden der empirischen Sozialforschung* (pp. 543–556). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-531-18939-0_38

Mayring, P., Fenzl, T., & Letz, F. (n. d.). *QCAmap // a software for Qualitative Content Analysis* [online]. Verein zur Förderung qualitativer Forschung – Association for Supporting Qualitative Research ASQ. Klagenfurt. www.qcamap.org

Meiches, B., & Hopkins, R. (2012). Regime Theory. In *Oxford Research Encyclopedia of International Studies.* Oxford University Press. https://doi.org/10.1093/acrefore/9780190846626.013.472

Mena, S., & Palazzo, G. (2012). Input and Output Legitimacy of Multi-Stakeholder Initiatives. *Business Ethics Quarterly*, *22*(3), 527–556. https://doi.org/10.5840/beq201222333

Merriam-Webster, Inc. (Ed.). (2022). *Merriam-Webster Dictionary*. https://www.merriam-webster.com/dictionary

Miles, E. L., Andresen, S., Carlin, E. M., Skjærseth, J. B., Underdal, A., & Wettestad, J. (2002). *Environmental regime effectiveness: Confronting theory with evidence*. *Global environmental accord*. The MIT Press. https://doi.org/10.7551/mitpress/2784.001.0001

MSI Integrity. (2020a). *Not Fit-for-Purpose: The Grand Experiment of Multi-Stakeholder Initiatives in Corporate Accountability, Human Rights and Global Governance.* MSI Integrity. https://www.msi-integrity.org/wp-content/uploads/2020/07/MSI_Not_Fit_For_Purpose_FORWEBSITE.FINAL_.pdf

MSI Integrity. (2020b). *Not Fit-for-Purpose: The Grand Experiment of Multi-Stakeholder Initiatives in Corporate Accountability, Human Rights and Global Governance (Summary Report).* MSI Integrity. https://www.msi-integrity.org/wp-content/uploads/2020/07/MSI_SUMMARY_REPORT.FORWEBSITE.FINAL_.pdf

MSI Integrity, & International Human Rights Clinic at Harvard Law School. (2017). *MSI Evaluation Tool: For the Evaluation of Multi-Stakeholder Initiatives*. https://www.msi-integrity.org/wp-content/uploads/2017/11/MSI_Evaluation_Tool_2017.pdf

Nonprofit Finance Fund, Center for Health Care Strategies, & Alliance for Strong Families and Communities. (2017). *Partnership Assessment Tool for Health*. https://nff.org/fundamental/partnership-assessment-tool-health

OECD. (1991). *Principles for Evaluation of Development Assistance*. Paris. https://www.oecd.org/development/evaluation/2755284.pdf

OECD. (2008). *Evaluating the effectiveness and efficiency of partnerships involving governments from OECD member countries* (ENV/EPOC(2006)15/FINAL). Paris. Environment Directorate.

OECD. (2019). *Better Criteria for Better Evaluation: Revised Evaluation Criteria Definitions and Principles for Use*. https://www.oecd.org/dac/evaluation/revised-evaluation-criteria-dec-2019.pdf

Oguntuase, O. J. (2020). Role of Transnational Multi-stakeholder Partnerships in Achieving Sustainable Development Goals. In W. Leal Filho, A. M. Azul, L. Brandli, A. Lange Salvia, & T. Wall (Eds.), *Encyclopedia of the UN Sustainable Development Goals. Partnerships for the Goals* (pp. 1–15). Springer International Publishing. https://doi.org/10.1007/978-3-319-71067-9_136-1

Olson, M. (1965). *The Logic of Collective Action. Public Goods and the Theory of Groups.* Harvard University Press.

Ostrom, E. (2010). Analyzing collective action. *Agricultural Economics*, *41*, 155–166. https://doi.org/10.1111/j.1574-0862.2010.00497.x

Ostrom, E. (2015). *Governing the Commons*. Cambridge University Press. https://doi.org/10.1017/CBO9781316423936

The Partnering Initiative. (n. d.). *Internal prospective partnership assessment*. https://www.thepartneringinitiative.org/wp-content/uploads/2018/12/Internal-prospective-partnership-assessment-tool.pdf

The Partnering Initiative. (2018). *Partnership health check*. https://www.thepartneringinitiative.org/wp-content/uploads/2018/12/Partnership-health-check.pdf

Pattberg, P. H., & Widerberg, O. (2016). Transnational multistakeholder partnerships for sustainable development: Conditions for success. *Ambio*, *45*(1), 42–51. https://doi.org/10.1007/s13280-015-0684-2

Prior, L. (2008). Repositioning Documents in Social Research. *Sociology*, *42*(5), 821–836. https://doi.org/10.1177/0038038508094564

Rasche, A. (2012). Global Policies and Local Practice: Loose and Tight Couplings in Multi-Stakeholder Initiatives. *Business Ethics Quarterly*, *22*(4), 679–708. https://doi.org/10.5840/beq201222444

Rasche, A., Waddock, S., & McIntosh, M. (2013). The United Nations Global Compact. *Business & Society*, *52*(1), 6–30. https://doi.org/10.1177/0007650312459999

Reynaers, A.-M., & Grimmelikhuijsen, S. (2015). Transparency in public-private partnerships: Not so bad after all? *Public Administration*, *93*(3), 609–626. https://doi.org/10.1111/padm.12142

Risse, T. (2004). *Transnational Governance and Legitimacy*. Paper presented at the Fifth Pan-European Conference on International Relations, The Hague.

Risse, T. (2012). Governance in Areas Of Limited Statehood. In D. Lēwî-Faur (Ed.), *The Oxford handbook of governance.* Oxford Univ. Press. https://doi.org/10.1093/oxfordhb/9780199560530.013.0049

Roberts, R. (2019). *The Role of Scientific Evidence in Collective Action Decision-Making*. Dissertation. ProQuest Dissertations Publishing.

Roger, C. B., Hale, T. N., & Andonova, L. B. (2017). The Comparative Politics of Transnational Climate Governance. *International Interactions*, *43*(1), 1–25. https://doi.org/10.1080/03050629.2017.1252248

Salheiser, A. (2014). Natürliche Daten: Dokumente. In N. Baur & J. Blasius (Eds.), *Handbuch Methoden der empirischen Sozialforschung* (pp. 813–827). Springer Fachmedien Wiesbaden.

Sandler, T. (2010). *Global Collective Action*. Cambridge University Press. https://doi.org/10.1017/CBO9780511617119

Scharpf, F. W. (1997). Economic integration, democracy and the welfare state. *Journal of European Public Policy*, *4*(1), 18–36. https://doi.org/10.1080/135017697344217

Scharpf, F. W. (2006). *Problem Solving Effectiveness and Democratic Accountability in the EU* (IHS Political Science Series Working Paper 107). https://irihs.ihs.ac.at/id/eprint/1692/1/pw_107.pdf

Scharpf, F. W. (2009). Legitimacy in the multilevel European polity. *European Political Science Review*, *1*(2), 173–204. https://doi.org/10.1017/S1755773909000204

Schmidt, W. (2017). Dokumentenanalyse in der Organisationsforschung. In S. Liebig, W. Matiaske, & S. Rosenbohm (Eds.), *Handbuch Empirische Organisationsforschung*

(pp. 443–466). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-08493-6_16

Stadtler, L. (2016). Scrutinizing Public–Private Partnerships for Development: Towards a Broad Evaluation Conception. *Journal of Business Ethics*, *135*(1), 71–86. https://doi.org/10.1007/s10551-015-2730-1

Steffek, J., Müller, M., & Behr, H. (2021). Terminological Entrepreneurs and Discursive Shifts in International Relations: How a Discipline Invented the "International Regime". *International Studies Review*, *23*(1), 30–58. https://doi.org/10.1093/isr/viaa003

Stufflebeam, D. (2001). Evaluation Models. *New Directions for Evaluation*, *2001*(89), 7. https://doi.org/10.1002/ev.3

Tsou, C., Haynes, E., Warner, W. D., Gray, G., & Thompson, S. C. (2015). An exploration of inter-organisational partnership assessment tools in the context of Australian Aboriginal-mainstream partnerships: A scoping review of the literature. *BMC Public Health*, *15*, 416. https://doi.org/10.1186/s12889-015-1537-4

Underdal, A. (1992). The Concept of Regime `Effectiveness'. *Cooperation and Conflict*, *27*(3), 227–240. https://doi.org/10.1177/0010836792027003001

Underdal, A. (2004). Methodological Challenges in the Study of Regime Effectiveness. In A. Underdal & O. R. Young (Eds.), *Regime Consequences* (pp. 27–48). Springer Netherlands. https://doi.org/10.1007/978-1-4020-2208-1_2

Underdal, A., & Young, O. R. (Eds.). (2004). *Regime Consequences*. Springer Netherlands. https://doi.org/10.1007/978-1-4020-2208-1

United Nations. *Partnerships for SDGs online platform*. https://sustainabledevelopment.un.org/partnerships/

United Nations. (1992). *Agenda 21*. https://sustainabledevelopment.un.org/content/documents/Agenda21.pdf

United Nations. (2002). *Draft plan of implementation of the World Summit on Sustainable Development: World Summit on Sustainable Development Johannesburg, South Africa 26 August–4 September 2002* (A/CONF.199/L.1).

United Nations. (2010). *Partnerships for Sustainable Development: Report of the Secretary-General* (E/CN.17/2010/13).

United Nations. (2012). *The future we want: Resolution adopted by the General Assembly on 27 July 2012.*

United Nations. (2014). *Global governance and global rules for development in the post-2015 era: Policy note*. United Nations publication.

United Nations. (2015a). *Partnerships for Sustainable Development Goals: A legacy review towards realizing the 2030 Agenda*. https://sustainabledevelopment.un.org/content/documents/2257Partnerships%20for%20SDGs%20-%20a%20review%20web.pdf

United Nations. (2015b). *Transforming our world: the 2030 Agenda for Sustainable Development*. https://sustainabledevelopment.un.org/content/documents/21252030%20Agenda%20for%20Sustainable%20Development%20web.pdf

United Nations. (2018). *Partnership Exchange 2018 Report*. https://sustainabledevelopment.un.org/partnerships/about

Ven, H. van der, Bernstein, S. F., & Hoffmann, M. J. (2017). Valuing the Contributions of Nonstate and Subnational Actors to Climate Governance. *Global Environmental Politics*, *17*(1), 1–20. https://doi.org/10.1162/GLEP_a_00387

Victorian Health Promotion Foundation. (2016). *The partnerships analysis tool: A resource for establishing, developing and maintaining partnerships for health promotion*. https://www.vichealth.vic.gov.au/-/media/ResourceCentre/PublicationsandResources/General/VH_Partnerships-Analysis-Tool_web.pdf?la=en&hash=695C103985C0FFA7C66C1C656014773AD6942347

WaterAid, & Tearfund. (2005). *An empty glass: The EU Water Initiative's contribution to the water and sanitation Millennium targets*. https://washmatters.wateraid.org/publications/an-empty-glass-discussion-paper-2005

Weimann, J., Brosig-Koch, J., Heinrich, T., Hennig-Schmidt, H., & Keser, C. (2019). Public good provision by large groups – the logic of collective action revisited. *European Economic Review*, *118*, 348–363. https://doi.org/10.1016/j.euroecorev.2019.05.019

Weiss, T. G., & Wilkinson, R. (2014). Rethinking Global Governance? Complexity, Authority, Power, Change. *International Studies Quarterly*, *58*(1), 207–215. https://doi.org/10.1111/isqu.12082

Westman, L., & Broto, V. C. (2018). Climate governance through partnerships: A study of 150 urban initiatives in China. *Global Environmental Change*, *50*, 212–221. https://doi.org/10.1016/j.gloenvcha.2018.04.008

Widerberg, O., & Pattberg, P. H. (2015). International Cooperative Initiatives in Global Climate Governance: Raising the Ambition Level or Delegitimizing the UNFCCC? *Global Policy*, *6*(1), 45–56. https://doi.org/10.1111/1758-5899.12184

Williams, M. S. (2009). Citizenship as Agency within Communities of Shared Fate. In S. F. Bernstein (Ed.), *Globalization and autonomy series. Unsettled legitimacy: Political community, power, and authority in a global era* (pp. 33–52). UBC Press.

Wolf, K. D. (July 2010). *Output, Outcome, Impact: Focusing the Analytical Lens for Evaluating the Success of Corporate Output, Outcome, Impact: Focusing the Analytical Lens for Evaluating the Success of Corporate Contributions to Peace-Building and Conflict Prevention* (PRIF Working Papers No. 3). Frankfurt am Main. Hessische Stiftung Friedens- und Konfliktforschung. https://nbn-resolving.org/urn:nbn:de:0168-ssoar-459512

Young, I. M. (2000). *Inclusion and democracy* (1st ed.). *Oxford political theory*. Oxford Univ. Press. http://www.loc.gov/catdir/enhancements/fy0636/00025844-d.html

Young, O. R. (1999). *Governance in World Affairs*. Cornell University Press. https://doi.org/10.7591/9781501711404

Zelli, F., Bäckstrand, K., Nasiritousi, N., Skovgaard, J., & Widerberg, O. (2020). *Governing the Climate-Energy Nexus: Challenges to Coherence, Legitimacy and Effectiveness*. Cambridge University Press. https://doi.org/10.1017/9781108676397

Zürn, M. (1998). *Regieren jenseits des Nationalstaates: Globalisierung und Denationalisierung als Chance* (1st ed.). *Edition zweite Moderne*. Suhrkamp.

Zürn, M. (2012). Global Governance as Multi-Level Governance. In D. Lēwî-Faur (Ed.), *The Oxford handbook of governance.* Oxford Univ. Press. https://doi.org/10.1093/oxfordhb/9780199560530.013.0051

# Appendix A

*Coding Guideline*

| No. | Category / Subcategory | Definition | Anchor examples | Coding rules |
|---|---|---|---|---|
| **1** | **Participation & Inclusiveness** | Refers to the participation and inclusion of stakeholders in the decision-making process, particularly with regard to stakeholders from the Global South, non-governmental organizations, and marginalized groups (cf. Bäckstrand, 2012; Dingwerth, 2007; Mena & Palazzo, 2012). | | |
| 1.1 | Participation & Inclusiveness_ internal | This subcategory refers to the participation & inclusion of external stakeholders which are not members of the MSI. | Does the membership of the partnership represent the right people?<br>Does it fully represent target groups?<br><br>Is there any particular stakeholder group which feels it has been excluded from the partnership? | Includes all questions regarding which stakeholder( group)s which are participating in the MSI and take part in its decision-making.<br>Does not include specifications about the decision-making process (cat. 5). |
| 1.2 | Participation & Inclusiveness_ external | This subcategory refers to the participation & inclusion of external stakeholders who are not members of the MSI. | Are users and the community involved in practice?<br><br>Did the MSI invite representatives specifically from affected populations (i.e., rights- holders) to participate in the formation process?<br><br>What mechanisms are in place to receive feedback from our target population? | Includes all questions referring to the participation and inclusion of external stakeholder( group)s into the decision-making of the MSI.<br>Does not include specifications about the decision-making process (cat. 5) nor the mere provision of information to external stakeholders (cat. 2.2). |

| | | | | |
|---|---|---|---|---|
| **2** | **Transparency** | Refers to the access to information about the work of the MSI, including available information as well as resources to access and process it (Bäckstrand & Kylsäter, 2014; Dingwerth, 2007). | | |
| 2.1 | Transparency_ internal | This subcategory refers to internal transparency, regarding information provided to MSI members about organizational structures and processes, decisions, project parameters, expectations, etc. | Does the partnership operate in a transparent way – for example, with clear understanding of the funding or other support contributed by each partner?<br><br>There are clear arrangements to ensure that monitoring and review findings are, or will be, widely shared and disseminated amongst the partners. | Includes aspects of internal transparency, regarding the available information for MSI members about organizational structures and processes, decisions, project parameters, expectations, etc. as well as the form and channels of its communication. |
| 2.2 | Transparency_ external | This subcategory refers to external transparency, regarding all information made available to the public or specific external stakeholder groups. | Is the purpose of the group known and understood outside the partnership itself?<br><br>Is adequate information available about the partnership and its decisions? | Includes aspects of external transparency, regarding the available information for external actors or the public about the MSI, its goals, structures, decisions, outputs, etc. MSI members about organizational structures and by which means and channels they are provided. |
| **3** | **Accountability & Control Mechanisms** | Refers to structures and practices which hold MSIs accountable for their actions, possible sanctions, control mechanisms and who carries them out and is able to access them (cf. Dingwerth, 2007; Grant & Keohane, 2005). | | |

| | | | | |
|---|---|---|---|---|
| 3.1 | Accountability & Control Mechanisms_ internal | This subchapter includes all accountability and control mechanisms within the MSI or emanating from the MSI. | There are clear lines of accountability for the performance of the partnership as a whole.<br><br>Is there a system for taking grievances alleging breach of the rules of internal governance?<br><br>Are evaluations used in the MSI framework to examine targeted actors' compliance with MSI standards? | Includes all questions regarding accountability and control mechanisms and practices which are in place within the MSI or by which the MSI can hold actors accountable who commit to its standards or certification regulations. |
| 3.2 | Accountability & Control Mechanisms_ external | This subcategory includes all accountability mechanisms and practices which are or can be performed by external actors towards the MSI and/or its members. | Does the MSI enable or encourage public enforcement of its standards in national and/or international processes? | Includes all questions regarding accountability and control mechanisms and practices by which external actors can hold the MSI and/or its members accountable. |
| **4** | **Responsiveness** | Includes all questions addressing whether an MSI acts according to the interest of the beneficiaries of its collective action effort and the actors affected by its decision (cf. Dingwerth, 2007). | By working together, how well are these partners able to respond to the needs and problems of the community?<br><br>Does the MSI comprehensively address the human rights issues that prompted its formation? | Includes all aspects addressing whether the MSI is orienting its actions towards the interests of the beneficiaries and actors affected by them and the underlying collective action problem.<br>Does not include questions whether an MSI is the best form to address it (cat. 6.9). |
| **5** | **Discourse & Deliberative Practice** | Refers to all aspects of the quality of discourse and deliberation processes such as consensus-orientation, power relations and participation in the process (Bäckstrand, 2012; Dingwerth, 2007; Mena & Palazzo, 2012). | Do partners feel they are able to influence decisions made by the partnership?<br><br>Conflicts of interest are managed effectively<br><br>Do both partners and other stakeholders feel their views are listened to, even when they may be critical? | Includes all questions regarding the practice of discourse, deliberation, and decision-making in the MSI, such as power-relations, consideration of different views, how decisions are being made, etc. Also entails aspects related to conflict resolution. |

| | | | | |
|---|---|---|---|---|
| **6** | **Institutional Effectiveness** | Refers to the adequacy of the institutional frame, structures, and design for achieving the outcome aspired by the MSI (Bäckstrand, 2006, 2012).<br>[Does not include aspects of institutional effectiveness which are part of categories 1-5.] | | |
| 6.1 | Institutional Effectiveness_ Vision & Purpose | This subcategory includes questions regarding the vision, mission, purpose of the MSI. | Does the group possess shared values and accepted principles?<br><br>Are the overall vision, purpose, and goals still recognized by members? | Includes all questions regarding the vision, mission & purpose as well as organization identity of the MSI.<br>Does not include goals and objectives which relate to specific output, outcome and impact targets and their measurement (cat. 7, 8, 9). |
| 6.2 | Institutional Effectiveness_ Roles & Responsibilities | This subcategory includes questions regarding the distribution of roles and responsibilities as well as ownership (of products, intellectual property, etc.) within the MSI. | Is there a clear understanding of own/other's roles and responsibilities?<br><br>Each partner's areas of responsibility are clear and understood. | Includes all aspects of how roles and responsibilities as well as ownership of products, intellectual property, etc. are organized and distributed within the MSI.<br>Does not cover which partners are included in the MSI in general (cat. 1) or ownership in the sense of commitment (cat. 6.6). |
| 6.3 | Institutional Effectiveness_ Resources | This subcategory includes aspects regarding the MSI's resources and their distribution and allocation. | Have the partners jointly discussed and formalized in an agreement what resources each partner will provide (financial, human resources, equipment, indigenous knowledge etc.)?<br><br>What is the level of satisfaction of the members of the partnership related to the allocation of adequate funds to support the dissemination plan? | Includes all questions regarding the resources of the MSI, who provides them and how they are allocated. Comprises financial as well as non-financial resources.<br>Does not include the distribution of financial revenues etc. to members (cat. 6.7). |

| | | | | |
|---|---|---|---|---|
| 6.4 | Institutional Effectiveness_ Communication | This subcategory includes aspects regarding the organization of communication of the MSI. | Do formal and informal communications take place?<br><br>The partners and leaders have good opportunity for face- to-face communication.<br><br>The lines of communication between the partners are clear. | Includes questions addressing the communication strategy and structures of the MSI, both internal and external publication and representation activities.<br>Excludes questions about the provision of information for transparency (cat. 2). |
| 6.5 | Institutional Effectiveness_ Ethics & Work Culture | This subcategory covers aspects regarding the working culture within the MSI and its ethical practice. | Has the partnership agreed a common code of conduct for its operations?<br><br>The partnership is planned to be culturally appropriate to the public/community/local setting.<br><br>The partners have an understanding and respect for each other. | Includes questions addressing the culture of working together within the MSI, regarding for example respect, trust, or recognition of other members' contribution. Also comprises the ethical practice of the MSI, both internally and externally.<br>Excludes the aspects of the discursive practice (cat. 5) and participation & inclusiveness (cat. 1). |
| 6.6 | Institutional Effectiveness_ Commitment | This subcategory covers questions regarding the commitment of the members to the MSI and its goals. | Are partners willing to make changes to achieve shared goals?<br><br>There is a clear commitment to partnership working from the most senior levels of each partnership organization. | Includes all questions addressing the commitment of MSI members to the MSI and its goals. Does also include ownership in the sense of commitment. |
| 6.7 | Institutional Effectiveness_ Benefits | This subcategory covers benefits for the MSI and/or its members as well as the motivation of members to participate in the MSI. Also includes possible drawbacks & risks. | Benefits derived from the partnership are fairly distributed among all partners.<br><br>What benefits do partners feel they have gained from the partnership, both individually and collectively? | Includes all questions regarding the benefits and value-added etc. as well as drawbacks for the MSI itself and/or its members. Also comprises the motivation and intention of members to participate in the MSI. |

| | | | | |
|---|---|---|---|---|
| 6.8 | Institutional Effectiveness_ Organizational Structures & Change | This subcategory includes aspects of organizational and governance structures of the MSI and their evaluation and adaptation over time. | There are clear arrangements effectively to monitor and review how the partnership itself is working.<br><br>Have any organizational improvements occurred after establishment of the partnership? | Includes questions regarding the organizational and governance structures of the MSI and their monitoring, evaluation, and adaptation to changed conditions.<br>Does not include communication structures (cat. 6.4). |
| 6.9 | Institutional Effectiveness_ Organizational Form | This subcategory covers questions addressing the aptitude of MSI as form of organization. | A partnership approach is the best way to address the issue at hand. | Includes questions regarding the form of organization in itself and its aptitude for the addressed collective action problem. |
| 6.10 | Institutional Effectiveness_ External Support | This subcategory covers aspects regarding networking and external support for the MSI. | The partnership is supported by policy leaders and influential decision-makers.<br><br>The partnership has the capacity to create strategic alliances and joint working arrangements across organizational boundaries. | Includes questions addressing the networking activities and external support for the MSI.<br>Does not include external stakeholder participating in the process (cat. 1.2). |
| 6.11 | Institutional Effectiveness_ Termination | This subcategory includes questions regarding the termination phase of the MSI and possible follow-ups. | Is there in place an exit (or 'moving on') strategy for the partnership?<br><br>Have the partners planned for project meetings and consultations to explore future directions, new relationship opportunities, transition possibilities and other concrete actions when the project ends? | Includes questions regarding the termination of the MSI, exit scenarios, the transfer of resulted learnings and possible follow-ups. |
| 6.12 | Institutional Effectiveness_ Other | This subcategory includes all aspects not covered by the previous subcategories of institutional effectiveness. | | Includes all aspects not covered by the previous subcategories of institutional effectiveness (cat. 6.1-6.11), for example regarding the identification of success factors for MSI working. |

| | | | | |
|---|---|---|---|---|
| **7** | **Output Effectiveness** | Refers to (self-)commitments of MSIs or MSI members regarding results, including certification standards or agreed principles (Jastram & Klingenberg, 2018; Wolf, 2010). | Has the group agreed upon a set of outcomes?<br><br>Does it monitor whether or not these objectives have been achieved? | Includes (self-)commitments of MSIs or MSI members in terms of aspired results, certification standards or agreed principles as well as their measurement and evaluation. It also includes the influence of results on policy making. |
| **8** | **Outcome Effectiveness** | Refers to changes in human behavior as a result of the aspired outputs (=self-commitments) made by the MSI or an MSI member (Jastram & Klingenberg, 2018; Wolf, 2010). | Have the members of the partnership jointly negotiated and agreed upon capacity building goals, including the roles and responsibilities of each member? | Includes changes in human behavior based on output results. Also comprises aspects of capacity building. |
| **9** | **Impact Effectiveness** | Refers to the contribution of the MSI attaining the Sustainable Development Goals and solving collective action problems (Jastram & Klingenberg, 2018; Wolf, 2010). | How is the partnership designed to assist both learners and teachers improve educational outcomes?<br><br>Overall, reviewing both quantitative changes such as retention rates and exam performance, and the perceptions of head teachers, teachers, parents, students, and community groups, what has been the partnership's impact on education in its local area? | Includes aspects of the effect of the MSI on the beneficiaries and the addressed collective action problem. |