CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Institute of Political Studies

Department of International Relations

Master's Thesis

2025   **Laura Mangunda**

CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Institute of Political Studies

Department of International Relations

Impact Evaluation Practices in Professionalized Development NGOs

Master's Thesis

Author of the Thesis: Laura Mangunda

Study programme: International Relations (MV)

Supervisor: doc. PhDr. Vít Střítecký, M.Phil., Ph.D.

Year of the defence: 2025

Declaration

1. I hereby declare that I have compiled this thesis using the listed literature and resources only.
2. I hereby declare that my thesis has not been used to gain any other academic title.
3. I fully agree to my work being used for study and scientific purposes.

In Prague on 7th January 2025   Laura Mangunda

References

MANGUNDA, Laura. *Impact Evaluation Practices in Professionalized Development NGOs.* Prague, 2025. 68 p. Master's thesis (Mgr). Charles University, Faculty of Social Sciences, Institute of Political Studies, Department of Inernational Relations. Supervisor doc. PhDr. Vít Střítecký, M.Phil., Ph.D.

Length of the Thesis: 152,532 characters with spaces

Abstract

On the backdrop of the expansion and professionalisation of international development NGOs in recent decades, this thesis explores how these organisations conceptualize, carry out, and ultimately struggle to measure the long-term impact of their work. Drawing on practice theory and critical debates about NGO-ization, the Washington Consensus, and the New Public Management, it uncovers a fundamental mismatch between donor-driven evaluation requirements and the ambitious definitions of "impact" outlined by the widely used OECD DAC criteria. It finds that although NGOs routinely conduct various monitoring and evaluation activities, only a tiny fraction of projects undergo rigorous *ex post* impact assessments. Resource constraints, short-term project cycles, and managerial preference for actionable insights deter more comprehensive studies. Through interviews and organizational case analyses, the study shows that NGOs largely assume impact without measuring it systematically, raising questions about who should be responsible for robust impact measurement. Ultimately, these findings highlight persistent tensions among donor expectations, operational constraints, and the pursuit of meaningful, sustainable development outcomes.

Abstrakt

V širším kontextu růstu a profesionalizace mezinárodních rozvojových nevládních organizací (NNO) v posledních několika dekádách zkoumá tato práce jak tyto organizace ve své práci konceptualizují, provádějí a v důsledku zápasí s měřením dlouhodobých dopadů (impaktů) své práce. S využitím *practice teory* a na základě kritických debat ohledně NGO-izace, Washingtonského konsenzu, a Nové veřejné správy práce odhaluje fundamentální nesoulad mezi donory vyžadovanými evaluacemi a ambiciózní definicí "impaktu" na základě široce používaných kritérií OECD DAC. Ukazuje přitom, že i když tyto organizace rutinně provádějí různé druhy monitorování a evaluace, pouze nepatrný zlomek projektů prochází důkladným následným hodnocením impaktu. Komplexnějším studiím brání omezené zdroje, krátkodobé projektové cykly a manažerská praxe preferující okamžitě aplikovatelné poznatky. Z rozhovorů a analýz vychází najevo, že NNO impakt do značné míry předpokládají, aniž by jej systematicky měřily, což vyvolává otázku, kdo by měl být za důkladné měření dopadu zodpovědný. Studie upozorňuje na přetrvávající napětí mezi očekáváními donorů, provozními omezeními a snahou o smysluplné, udržitelné výsledky

rozvojových intervencí.

Keywords

NGO, Non-Governmental Organisations, impact, impact evaluation, practice theory, development, development discourse, evaluation, impact assessment, OECD DAC

Klíčová slova

neziskové organizace, impakt, hodnocení impaktu, nevládní organizace, rozvojová pomoc, evaluace, OECD DAC, hodnocení dopadu

Title

Impact Evaluation Practices in Professionalized Development NGOs

Název práce

Praktiky hodnocení impaktu v profesionalizovaných rozvojových nevládních organizacích

Acknowledgement

Table of Contents

Introduction

A foreign policy tool and simultaneously a tool for global poverty alleviation, international development has been a dynamic and significant part of modern international politics. With nearly half a century of evolution and ballooning within this industry, the impact assessment of international development NGOs, which are the preferred tool of aid delivery, is as relevant as ever.

Over the past decades, the funding and implementing structures of international development have undergone significant changes: Once primarily state-implemented efforts in post-war and (post-)colonial contexts, with the globalization and privatization trends from the 1980s onwards, an increasing portion of bilateral and multilateral development aid has been channeled through specialized agencies to private entities, including non-governmental organizations (NGOs). Today, most international development NGOs get the majority of their funding in competitive grants from institutional donors in so-called project cycles that last anywhere between several months to several years.

This shift has had a profound effect on the organizational structure and functioning of these NGOs, evident in their growth and professionalization over the past three decades. This process has led to a paradox, whereas NGOs that came to be preferred in aid delivery due to a comparative advantage in their intimate knowledge of the beneficiaries and of the situation "on the ground," their ability to informally network with a wide range of actors, and their flexibly to adapt to changing circumstances, came to be constrained by bureaucratic requirements and standards similar to those that previously limited the state-implemented efforts.

The subsequent scholarly debate about NGOs as actors not only in international development but also generally in international relations has generated many works concerning the issues of their character, legitimacy, effectiveness, efficiency, autonomy, and accountability. Aside from NGOs, the very concept of development and developmental aid has been criticized for its underlying assumptions and partial objectives rooted in Eurocentrism and for its failure to address the structural reasons for poverty, marginalization, and exclusion–and therefore failure to bring about real, lasting solutions.

This thesis builds on these debates and key concepts of the professionalization of NGOs, including the New Policy Agenda (NPA) and the neoliberal Washington Consensus, dynamics of donor-NGO relations, NGO-ization, and the nature of the New Public Management (NPM). It accounts for critical reflection on this process, embodied in issues of mission drift, depoliticization, contract culture, and multiple accountabilities, but also generally in the critique of development discourse. Building upon these debates, it aims to analyze how development NGOs that strive for high-level impacts but operate within the constraints of the aid system understand and evaluate the impact of their work. Following up on the recent linguistic turn and practice turn in social sciences, it uses practice theory as a tool to uncover the day-to-day practices guiding the process of impact evaluation of the

observed international development NGOs, and especially in an attempt to reconstruct the implicit, background knowledge guiding these practices.

Building on the theoretical background presented in earlier chapters, two key research questions guide this inquiry: first, how professionalized NGOs evaluate the impact of their work (RQ1), and second, what background knowledge underlies these evaluation practices (RQ2). A sample of three medium-sized, European international development NGOs—People in Need, Helvetas, and Concern Worldwide—was selected for their shared characteristics as professionalized, Western-based organizations and their institutional contexts and practices.

The sociopolitical relevance of this research is particularly evident in view of the considerable volume of development grants awarded annually by bilateral and multilateral institutional donors. Ongoing theoretical and practical debates on the legitimacy and effectiveness of international development in general—and NGOs as key actors in particular—further add to the perceived relevance.

Ever since NGOs emerged as a preferred means of delivering international development aid, their budgets, as well as their global influence, have expanded substantially. These organizations continue to portray themselves as independent entities with closer ties to local realities and stronger capacities to identify and implement meaningful interventions for those in need. They also function as a normative force in shaping international agendas, prominently showcasing the intended impacts of their work to both the public and the broader community. Consequently, understanding the practice of measuring this impact is of critical significance.

The results of this study begin by outlining the current approaches to impact evaluation within the observed NGOs, followed by an exploration of the specific obstacles these organizations encounter in such evaluative practices. Chief among the conclusions is that, despite a clear interest in assessing impact, NGOs seldom measure it in a manner consistent with its most widely accepted definitions. Several practical constraints inhibit these organizations from engaging in fully fledged impact assessments, such as the methodological complexity of the undertaking, the disproportionately high costs (in terms of both financial and human resources), the short-term project cycles on which NGOs remain dependent, and the limited applicability of the resulting data. Additional findings point to the problematic reliance on OSCE DAC indicators as a universal standard, the ambiguity surrounding the very definition of impact, managerial preference for immediately actionable insights, and, more broadly, the question of whether measuring impact ought to be an organizational priority in the first place.

Several avenues for further research were identified. One promising direction is the reconstruction of tacit knowledge regarding donor policies and pressure perceptions, or uncovering such knowledge in instances of conflict between stakeholder and donor interests. In general, the subject of impact assessment in development NGOs provides fertile ground for continued investigation, especially amid prospective shifts—such as donor bodies

extending project timeframes to include full impact assessments, evolving perceptions of NGOs as increasingly politicized actors (and potential protective measures against their operations), and the transformation of Western-based NGOs as they incorporate more robust local stakeholder inputs. At the same time, the rapidly changing development space, driven in large part by the escalating climate emergency, ensures that questions concerning the evaluation of NGO impact will remain highly relevant well into the future.

1. Theoretical background and key concepts

1.1. Introduction

A foreign policy tool and, simultaneously, a tool for global poverty alleviation, international development has been a dynamic and significant part of modern international politics. This chapter begins by defining NGOs as distinctive actors in the international space. It provides an overview of their development from largely understudied and marginal actors, through their growth in prominence and establishment within international law and expansion of their agenda, into the dynamic development in the last four decades, when NGOs became a powerful actor in shaping international norms and delivering development and humanitarian aid. It also provides an overview of how the dominant international relations (IR) theories approach NGOs as actors and the degree of relevance each school attributes to the non-governmental sector.

Furthermore, it lays down the process of professionalization and bureaucratization of NGOs since the 1980s, including the background provisions and developments that motivated NGOs into this process. These processes lay down the foundation for later chapters describing key concepts relevant for the scope of this research. A description of the specific processes that embody the process of professionalization—including the New Policy Agenda (NPA) and the neoliberal Washington Consensus, dynamics of donor-NGO relations, and the nature of the New Public Management (NPM)—are complemented by critical reflections on the process of professionalization and the domination of bureaucratic practices in NGO functioning, including issues of mission drift, depoliticization, contract culture, and multiple accountabilities.

Continuing with the overview of the key concepts framing the aim of this study, development discourse is outlined, highlighting how knowledge and power converge to shape ideas about social change. Building on the writings of Foucault and post-structuralists, especially on how language systematically organizes power through specific subjectivities and actions, the rise of NGOs as development actors, particularly in the 1980s, is examined, revealing how discursive frameworks legitimize or marginalize certain interventions. Key topics of upward versus downward accountability, donor preferences over local needs, external "expertise" overshadowing community perspectives, and other related topics are discussed. The notion of development discourse is outlined, offering a critical lens for addressing the two research questions, emphasizing how impact evaluation practices are both influenced by and help reproduce dominant narratives and implicit assumptions about progress and transformation.

Following is the concept of impact evaluation practice, advancing closer towards the practical considerations of the focus of this study. Whereas previous concepts are largely academic, the impact evaluation practice represents a practical and specific framework. While outlining the approach of NGOs in assessing the impacts of their interventions, it highlights systemic challenges like ambiguous definitions (e.g., "impact" vs. "outcome"), the limitations of linear causality, and debates over attribution. The OECD DAC Criteria and World Bank guidelines highlight a diversity of approaches—ranging from randomized controlled trials (RCTs) to theory of change models—yet reveal inconsistencies in

application and donor-centric biases. The micro-macro paradox further illustrates the tension between successful project-level assessments and underwhelming aggregate results at a national scale. This concept foregrounds the complexities NGOs face in justifying and measuring "impact" and offers a theoretical basis of best practices and common approaches that serve as a valuable tool in the practical section.

Lastly, this chapter outlines the concept of practice in international relations through the lens of practice theory, a key methodological lens for this research. This approach highlights how everyday actions—rather than abstract structures—shape and are shaped by social contexts. Originating from the works of Ortner, Bourdieu, and others, and later refined by Adler and Pouliot for international relations, the theory emphasizes how practices are repeated, socially meaningful performances underpinned by shared background knowledge. This background knowledge, often tacit, guides practitioners' interpretations and behaviors, enabling both the stabilization and transformation of institutional norms. Practice theory also highlights how "communities of practice," composed of individuals and organizations, sustain and transmit these shared understandings over time. By distinguishing between behavior, action, and practice, the framework uncovers the deeper cultural and discursive dimensions behind seemingly technical routines.

The guiding aim informing the structure of this chapter is to go from broader overviews framing this research topic in the vast study of international relations and global politics, gradually towards more specific and practical concepts. This way, hopefully, it directs the reader towards some key considerations and findings that underpin the selection of the topic of this study and the rationale for using its methodological approaches, which are further elaborated in the following chapter.

1.2. NGOs as distinct actors in IR

1.2.1. Defining NGOs

The definition of what an NGO is and what its defining factors are is influenced by the analytical perspective employed and the perceived importance of them as actors, or the degree of independence they have. Most broadly, NGOs are defined by the United Nations (UN) as "any international organization which is not established by inter-governmental agreement" (Ahmed & Potter 2013). However, such a definition is too broad to be useful in this context. More specifically, NGOs are defined as non-state, non-profit organizations that are often engaging in various forms of advocacy and service provision on local, national, or international levels in fields like human rights, development, and environmental protection to address and support the public good (ECOSOC 1996; Reus-Smit & Snidal 2008; Carlsnaes, Risse & Simmons 2002). Ahmed and Potter (2013) also use a helpful negative definition, stating that an NGO cannot be profit-making; it cannot advocate the use of violence; it cannot be a school, a university, or a political party; and any concern with human rights must be general rather than restricted to a particular communal group, nationality, or country.

NGOs can be differentiated based on the scope of their work as (1) *domestic or local NGOs* that operate within a specific community or region, focusing on local issues; (2) *national NGOs* that function within a single country, addressing national issues; and (3) *international NGOs* (INGOs) that work across borders, often tackling global challenges (Paul, Israel & Weltbank 1991; UN 2003; Reus-Smit & Snidal 2008; Carlsnaes, Risse & Simmons 2002). For the purposes of this research, NGOs are only approached as international actors, working across borders, operating in the international sphere, and engaging various local communities, states, international bodies, and other actors[1].

NGOs can also be classified based on the topical scope, purpose or focus of their work, mainly into: (1) *development NGOs* that focus on social and economic development, often working on issues like poverty alleviation, education, and health; (2) *humanitarian NGOs* that provide emergency relief in crisis situations such as natural disasters, armed conflicts or displacement; (3) *environmental NGOs* that advocate for the protection and preservation of the environment and natural resources; (4) *human rights NGOs* promoting and protecting human rights, advocating for justice, freedom, and equality; (5) *advocacy NGOs* that focus on influencing policies and public opinion on specific issues, such as women's rights, climate change, or labor laws; and (6) *operational NGOs* that primarily engage in project implementation, offering services or running programs directly in the field (Reus-Smit & Snidal 2008; Carlsnaes, Risse & Simmons 2002). While there are many examples of single-purpose NGOs, often an NGO will engage in more than one sphere, sometimes blurring the lines between one or the other. This research focuses primarily on development NGOs, but, acknowledging that the developments in one area will often influence the functioning of the entire organization and the effects of spillover, it will discuss developments in other spheres relevant to the scope of this research.

Based on the origin, most broadly, NGOs are separated into Northern NGOs (NNGOs), based in the global North, and Southern NGOs (SNGOs), from the global South. This research focuses on the functioning of NNGOs and any mention of "NGOs" is meant to refer primarily to those.

Another differentiator of NGOs can be their size according to various metrics, including the proportion of the market they control, operational spending, number of staff, global reach, and impact. During the 1990s, the literature talks about the "Big Eight" NGOs—the largest organizations that control about 50% of the relief market (Simmons 1998). Since then, the situation has evolved, but several of the largest NGOs still control a significant portion of the NGO market, including organizations such as the International Federation of Red Cross, CARE, World Vision International, Oxfam Federation, Doctors Without Borders, and Save the Children.

Without delving into the specifics, today, large NGOs can be broadly defined as operating with a budget of over 500 billion US dollars and a staff of tens of thousands. Based on this,

---

[1] Therefore, any later mentions of NGOs inherently refference INGOs only.

there are about 5 to 10 large NGOs in the global North. These household-name organizations are often mentioned together in reports by international organizations like the United Nations, the World Bank, and the European Union and in academic studies (Simmons 1998, Lewis, Kanji, Themudo 2021). Medium-sized NGOs operate with a budget of approximately several hundred million US dollars and a staff of thousands. Based on that definition, there are approximately 50 to 100 medium-sized development NGOs[2] in the global 'North' (Banks, Hulme 2012, Concern Worldwide 2024, Helvetas 2024, People in Need 2024).

1.2.2. Evolution of the role of NGOs in international relations

NGOs in various forms have been shaping global norms and humanitarian values and influencing international relations for decades. The conventional view in the study of NGOs in IR focuses on their emergence in the early 20th century, but some authors trace their influence well into the 18th and 19th centuries and beyond (Davies 2014). Reflecting the sentiments in international relations at that time, in the early 20th century and the interwar period, NGOs focused on peace promotion and humanitarian efforts, reacting to the horror of the First World War (*ibid.*). The end of the First World War marked a shift toward recognizing NGOs as distinct entities in the international system (Risse 2013). The creation of the United Nations (UN) post-Second World War and the inclusion of Article 71 in the UN Charter formalized the status of NGOs, marking the beginning of the role of NGOs as intermediaries between state actors, local communities, and international organizations like the UN (Simmons and Martin 2002). Thus, marking them as a formal part of the international system (Risse 2013).

Crucially for the scope of this research, the 1940s and 1950s saw the diversification of NGO activities beyond peace and humanitarianism to include issues like development, gender equality, and public health, deepening their professionalization and specialization along the way (Davies 2014). Davies highlights that the 1960s and 1970s were dominated by the formation of transnational advocacy networks, where NGOs across different countries collaborated to influence global governance.

The 1980s and 1990s were an era of rapid expansion of the NGO sector, with the number of development NGOs in the global 'North' and their funding nearly doubling (Smillie et al. 1993; Gross Stein 2011). As Smillie et al. (1993) point out, along with this development rose the scope of their work, with provisions of health, education, and credit services to millions of people. NGOs at this time reacted to the neoliberal turn in international relations, more deeply focusing on development and poverty reduction and mitigating the impacts of market solutions that dominated the thinking at this time (Davies 2014). The rise of the number of NGOs in the global 'North' at this time was also influenced by the decolonization of Soviet Socialist Republics and the breakup of Yugoslavia, coinciding with the disintegration of the relationship of former communist satellites with the USSR (Waller 1993).

---

[2] Practically, in order to match the size of the larger NGOs, medium and small-sized organizations often enter into coalitions.

The 1980s and particularly the 1990s meant a period of reflection and changing of the *modus operandi* of international humanitarian and development NGOs in reaction to the rise of a new type of conflict—the intrastate war with unclear conflict parties on the backdrop of weak state sovereignty, blurring the lines between war, crime, and political violence (Themnér & Wallensteen 2012; Kaldor 2003). Fatal shortcomings in the practice of international humanitarian NGOs at this time, particularly in the 1994 genocide in Rwanda, coupled with increased pressure from state donors providing all the extra funding, were the catalyst to a process of further professionalization and development of monitoring practices in NGOs (Gross Stein 2011). Some authors attribute the rise of the global civil society in the 1990s also to the failure of traditional state actors to address new forms of conflict and crises (Kaldor 2003).

Post-2000, NGOs are increasingly integrated into the global governance structures and engage with state and non-state actors and international organizations, playing a key role in norm diffusion and policy advocacy, particularly in human rights, democratization, and environmental protection (Risse 2013; Edwards & Hulme 2014). This era also sees the continuation of the growth of the number of NGOs and their resources, as governments see them as a cost-effective way to deliver welfare and assistance (Edwards & Hulme 2014). With increasing pressures on accountability (both towards donors and local communities), NGOs are increasingly pressed towards professionalization, including practices like regular audits, staff training, and comprehensive reporting (*ibid.*).

During the 1990s, NGOs still enjoyed relative autonomy alongside their growing influence over global governance. Authors like Davies (2014) track the process of NGO-ization and state co-option, especially after the turn of the millennium. This development is seen as a backlash against the perceived weakening of the role of the state in the 1990s in favor of international NGOs and is embodied in the growth of mechanisms through which the state exerted influence over NGOs–including financial dependency, accompanied by reporting and conditionality, the establishment of state-controlled NGOs (referred to as Government-Organized Non-Governmental Organizations or GONGOs) or introducing legal restrictions, such as requiring compulsory registration for NGOs (*ibid.*).

1.2.3. Theoretical approaches to the role of NGOs in international relations

Different theoretical approaches to international relations see the role and significance of NGOs differently. To set the baseline and position this research within the broader debates within IR, some space needs to be dedicated to an overview of the main IR theories and their approaches to NGOs. According to Ahmed and Potter (2013), there has been relatively scarce attention given to how NGOs fit into the broader IR theory because their study often crosses disciplinary and theoretical boundaries, and the emergence of the influence of NGOs on international relations has been quite recent. Moreover, NGOs' growth in prominence can be traced only to the last four decades, when their funding and numbers proliferated.

Classical realists are skeptical about the influence of NGOs, as they see international relations as an anarchic system dominated by the struggle for power by states as the main actors (Wohlforth 2008; Krasner 1992; Ahmed & Potter 2013). For these reasons, they see NGOs as peripheral actors that are primarily tools for states to advance their national interests, without fundamentally altering the distribution of power within the system (*ibid.*). The same goes for neorealism (structural realism), which further marginalizes the role of NGOs in the international system (Wohlforth 2008). Kenneth Waltz, the godfather of neorealist thought, clearly states that although states are not the only international actors, the structure of the international system is defined not by all actors, but only by the major ones, thus, by states (Waltz 1986).

Liberalism and neoliberal institutionalism focus on a more peaceful world dominated by cooperation and, thus, see NGOs more favorably, recognizing them as key actors that promote cooperation and facilitate international governance (Martin & Simmons 2013; Moravcsik 2008). From a liberal perspective, NGOs are key players in the international system because they facilitate cooperation between states and promote norms of governance, human rights, and environmental protection, along with technical expertise in key areas (Martin & Simmons 2013). They are seen as vital in fostering dialogue, reducing transaction costs, and contributing to the global governance architecture to tackle transnational challenges like climate change, health crises, and humanitarian disasters (*ibid.*). Liberals see NGOs working alongside other actors in the international space—multinational corporations (MNCs), sub-national actors, agencies, or international organizations (Ahmed & Potter 2013). However, the emergence of the interdependence theory focused more on MNCs and the subsequent neo-neo debate still failed to resolve whether the state remains the major object of study (*ibid.*).

As an outgrowth of neoliberal institutionalism, regime theory focuses more deeply on NGOs, as they highlight informal interactions and interdependence in the creation of international regimes that govern the global community (Keohane 2005; Martin & Simmons 2013). Regimes typically consist of state governments and international laws, but research on the emergence and sustenance of these regimes repeatedly points out the contributions of non-state actors, including NGOs (Ahmed & Potter 2013). NGOs, alongside interest groups, transnational coalitions and individuals, approach governments to solve international problems, which can then be established into a legal international framework, such as some standards of humanitarian or environmental protection (*ibid.*).

Regime theory offers a bridge towards constructivism, as some constructivist scholars contributed to the evolution of regime theory (Adler 2013; Martin & Simmons 2013). However, constructivism focuses more on the ideational process within international relations, highlighting the role of norms, ideas, and social constructs in shaping state behaviour (*ibid.*) Constructivists see regimes not only as pragmatic frameworks for cooperation but as key vehicles for shaping the identities and interests of states through shared norms and values (Wendt 1999; Adler 2013). Within the constructivist debate, international NGOs are often perceived as key creators of norms and values that are

integrated into international practices, giving them great power in shaping the international system (Risse 2013). Given this 'entrepreneurial' power, NGOs often drive the agenda in areas where state action is lacking, having the power to transform state identity on issues of future importance.

Neo-Marxist theories, especially the world-systems theory and dependency theory, generally view NGOs with skepticism, highlighting their role in the structures of global capitalism, arguing that NGOs present themselves as advocates of the marginalized but may, in fact, reinforce the global capitalist system of global inequality instead of addressing its root causes (Hönke & Lederer 2013). They might provide humanitarian or development assistance but often operate within a framework that perpetuates the exploitation of the global South by the global North (*ibid.*). Neo-Marxists focus on the way NGOs might be co-opted by corporate and state interests, thus becoming part of the machinery that sustains global economic hierarchies rather than challenging them (Cox 1996). Other authors criticizing the neoliberal global order see NGOs as part of the process of depoliticizing economic and social issues by framing them as technical or humanitarian problems rather than issues of systemic inequality and focus on the reduction of state power in favor of corporations, including through NGOs substituting state services (Harvey 2020).

However, not all critical authors see the role of NGOs in the international system negatively. Some authors from the Frankfurt school argue that NGOs have the potential to challenge existing power structures by fostering emancipation and promoting global justice (Zehfuss 2013). But overall, critical theorists focus on the critique of the role of NGOs. Post-structuralists see them through the lens of power discourse, contributing to the production of discourses that define "development," "human rights," or "security," thus shaping how global problems are understood and addressed and reinforcing the Western-centric discourses (*ibid.*). Postcolonialists like Edward Said, Gayatri Spivak, and Dipesh Chakrabarty argue that NGOs often operate within a framework of neo-imperialism, enforcing Western-centric values on the global South (Zehfuss 2013; Hönke & Lederer 2013).

Feminist authors such as J. Ann Tickner argue that NGOs are vital in advocating for women's rights and gender equality in international relations (Tickner 2014). They acknowledge the critical role NGOs, particularly women's organizations, play in addressing issues like sexual violence in conflict zones but also critique NGOs for not doing enough to challenge the patriarchal structures within which they operate and for imposing Western feminist frameworks that may not align with the needs and values of women in the global 'South' (*ibid.*).

1.3. Development of NGOs: Growth and professionalization

1.3.1. Conditions for the growth and professionalization of development NGOs

As illustrated in the chronological evolution of the NGOs as actors in the international space and their greater prominence in international relations theories that grew in popularity from the 1980s onwards, the last four decades saw the greatest rise of influence of NGOs within international relations. Indeed, the 1980s and 1990s saw an explosion in the number of development NGOs—within the OECD and the global North, their number grew from 1,600 in 1980 to 2,970 in 1993 (Smillie et al. 1993). Official OECD documents from the year 2000 estimate at least 3,900 development NGOs in Europe alone, far exceeding the previous OECD estimates and signifying the sector's growth in these years (Woods 2000).

The same development can be traced to their financing: the total spending of development NGOs in the OECD countries rose from 2.8 billion USD in 1980 to 5.7 billion USD in 1993 (OECD 1994). This growth in official development spending channeled through NGOs has been coming from various sources, including international organizations (IOs) such as the World Bank and the EC/EU institutions (Edwards & Hulme 1996), or directly from individual country donors consistently led by the United States, Germany, the United Kingdom, Japan, and France (OECD 2024). While no current aggregate numbers are readily available, current numbers are referenced broadly to be in the realm of "tens of billions" in total annual NGO development spending among OECD countries[3].

Current research highlights the growth in importance of development NGOs in the last four decades and suggests that the degree of contribution of development NGOs might have been underestimated and that they play an even greater role than previously anticipated. A 2015 study of British-based development NGOs found that the development NGO sector spent nearly 7 billion GBP, which is about half of the UK government's official development assistance (ODA) that year (Banks & Brockington 2019). Davis points out that NGOs in development aid expanded to the point where, in 2015, NGOs expended 35 billion USD in ODA to eligible countries, exceeding individual bilateral aid from the majority of Development Assistance Committee (DAC) countries. Overall, ODA channelled through NGOs reached around 16 percent of total bilateral ODA, resulting in a push away from state-based development systems towards NGOs (Davis 2019).

1.3.2. NGO Growth and Professionalization: The Process and Critical Reflections

Between 1990 and 2023, NGOs underwent a significant process of professionalization and bureaucratization, driven by increasing demands for accountability, transparency, and efficiency from both donors and stakeholders. The growing influx of public and private

---

[3] This estimate is based on the usual percentage of government ODA delivered via NGOs in OECD in the range of 15-25%, depending on the country, combined with the spending of large development NGOs in the range of 1-2 billion per year, and mid-sized NGOs in the range of tens to hundreds of millions multiplied by their numbers (World Vision International 2023; Save the Children International. 2023; Oxfam 2021; Médecins Sans Frontières 2024, Concern Worldwide 2024, Helvetas 2024, People in Need 2024, Morton 2013)

funds, particularly from state agencies and international organizations, required NGOs to adopt more formalized structures and practices to ensure the responsible use of resources. This shift led to the integration of business-like management techniques and organizational standards, transforming the traditionally grassroots, volunteer-driven sector into one characterized by managerial expertise and performance metrics. As Brière et al. highlighted, project managers in NGOs faced growing expectations to demonstrate competencies similar to those found in private sector counterparts, reflecting the sector's shift towards a more professionalized workforce focused on delivering measurable outcomes in increasingly complex environments (Brière et al 2015; Fowler 2011).

The professionalization of NGOs as a result of external donor pressure has also meant the adoption of formalized reporting, planning, and program evaluations. NGOs responded to this pressure by introducing independent financial audits, quantitative program evaluations, and formal reporting mechanisms. According to Hwang and Powell (2009), NGOs became more bureaucratized, with increased reliance on strategic planning and measurable outcomes, moving away from informal practices towards more standardized and accountable operational models. This transformation resulted in the development of consistent management processes, enabling NGOs to align their missions with donor priorities, but at the cost of potential mission drift, as they prioritized short-term, quantifiable successes over long-term systemic change (Hwang & Powell 2009; Maier et al. 2016).

Golini et al. (2015) observed that NGOs dealing with international development projects progressively adopted project management tools and frameworks, initially designed for the corporate sector, to improve project success rates and accountability. Crack illustrates this on their adoption of the New Public Management (NPM) designed for the public service (Crack 2019). This adoption of formal management techniques, including risk management, stakeholder analysis, and logical frameworks, enabled NGOs to standardize their practices across diverse operational contexts, ensuring that projects were both effective and transparent. However, as NGOs became more reliant on these tools, they also faced the risk of reducing complex social issues into technical problems, thereby limiting their ability to address the root causes of poverty and inequality (Golini et al 2015; Hwang & Powell 2009).

The professionalization and bureaucratization of NGOs during this period can also be understood within the broader context of sectoral blurring, where the boundaries between nonprofit, business, and government organizations have increasingly converged. Bromley and Meyer (2014) argue that this blurring is rooted in the global cultural shift towards rationalization and the application of science-like principles, which emphasize efficiency, accountability and formal structures across all sectors. NGOs thus adopted these practices not only due to donor pressure but also as part of a broader societal expectation that organizations—regardless of their sector—operate in a methodical and accountable manner. This trend has led to the rise of hybrid organizations that combine elements from all three sectors, further accelerating the professionalization of NGOs and embedding business-like practices into their operations (*ibid.*).

The approach towards NGOs from the 1990s onwards was summed up by a term that made the title of the seminal work of Edwards and Hulme on this topic—as a 'magic bullet' which can be fired off in any direction and, though often without very much evidence, will still find

its target; and the organizations themselves as a 'favored child' of official agencies (Edwards & Hulme 2014; Vivian 1994). Governments and international donors turned to NGOs as flexible, adaptable actors capable of delivering aid more efficiently than state mechanisms.

The individual donor agencies' approaches vary, but the overreaching trend has been governed by what researchers in the field call the New Policy Agenda (NPA) in NGO-state relations (Robinson 1994; Edwards & Hulme 1996). This approach emphasized market solutions and good governance and NGOs were seen as ideal partners for implementing NPA. Authors describe NPA as the development of closer relationships between governments and NGOs, increasing the reliance of NGOs on state funding and acting as service providers, implementing projects aligned with state or IO donor priorities, rather than independent agents of change. The impact of conditional state funding was later accompanied by the increased co-option of NGOs into the corporate world and their perceived service of the corporate interests of MNCs through multi-stakeholder initiatives and corporate social responsibility (CSR) programs (Davies 2014). Davies further points to the limiting of the radical potential of NGOs and their capacity to offer genuine alternatives to neoliberal globalization as a result of this process during the early 2000s (*ibid.*).

Edwards and Hulme identify two key elements of this policy. In terms of economy, especially in the 1990s and influenced by the neoliberal approach, governments started to favor market solutions and NGOs as the preferred way to achieve economic growth, producing goods and providing services, adhering to the mantra that "imperfect markets are better than imperfect states" (Edwards & Hulme 1996). The authors warn that this shift occurred without any convincing evidence of the superior performance of NGOs in this practice. Politically, the 1990s and early 2000s saw the dominance of 'democracy' as the gold standard in international relations[4]. Edwards and Hulme describe how good (*meaning 'democratic'*) governance began to be seen as essential for a healthy economy, even though the evidence for this relationship is mixed at best, and donors began to award key roles in democratization to NGOs with NPE (*ibid.*).

In association with the process of growing dependence of NGOs on state funding, authors (as back as from the 1980s (see Jan Prong in Hellinger, Hellinger & O'Regan 1988) are warning of the possibility of corruption of the missions of NGOs and deviation from their mission for social transformation in favor agendas of their donors (Edwards, Hulme 2014). To describe the process of growing dependence on donor funding, accompanied by professionalization, institutionalization and depoliticizing of their mission, authors began using the term NGO-ization (Hearn 1998; Sadoun 2006). Although this term is rooted in activist circles, describing the process of institutionalization of social movements, particularly women's movements, turning them away from their grassroot origins (see Batliwala 2007; Roy 2014) and in the context of NGOs in the global 'South' (Chahim &

---

[4] Signified by the dominance of the 'fight for democracy' in US foreign policy at that time, and the adoption of 'democratic' conditionality by the UN, for example in the Standards before Status condition for the recognition of Kosovo (Hill 2022, Thompson 2010).

Prakash 2014; Jad 2007), in the context of development NGOs based in the global 'North', authors focus on the process of growing dependence of development NGOs on state and IO funding, 'depolitization[5]' of their work in favor of a more technocratic provision of services, or the transformation of complex issues into short-term projects with quantifiable outcomes.

Other authors focus on the so-called 'contract culture' in financing NGOs under NPA, where the traditional matching grant model—where NGOs received funding to match their own resources—gradually gave way to a *contractual* model—where NGOs had to compete for government contracts to deliver services within predetermined frameworks—leading to professionalization, but also bureaucratic organizational culture (Smillie et al. 1993). Ossewaarde et al. (2008) further investigate this process and conclude that it then leads to a relationship that resembles the dynamics between a contractor and client. These contracts are performance-based, requiring NGOs to deliver specific, predefined services and outcomes within tightly controlled frameworks and Ossewaarde et al. warn that this imposes a market-oriented approach on NGOs, transforming them from independent actors advocating for social change into service providers that must prioritize efficiency, cost-effectiveness (*ibid.*).

Financial pressures from donors were not the only motivating factor that pushed NGOs towards professionalization of their practices at this time. Janice Gross Stein discusses how humanitarian NGOs have undergone a process of professionalization also in response to ethical concerns and criticisms surrounding their practices, particularly after the Rwandan genocide (Gross Stein 2011). Stein describes how these organizations, which had previously operated with a taken-for-granted sense of moral legitimacy based on "good intentions," began to face significant pressure from within and outside their communities to reflect critically on their actions. Influential voices like Mary Anderson and Alex de Waal highlighted how humanitarian efforts, even when well-intentioned, sometimes caused unintended harm by perpetuating conflicts or sustaining harmful political actors. This realization led to intense debates within organizations like Médecins Sans Frontières (MSF), CARE, and Oxfam over the need to formalize standards, improve accountability, and adopt practices that "do no harm." This process illustrates that the process of professionalization, although primarily driven by donor pressures, was also motivated by internal reflection and arose from a deeper ethical imperative to ensure that their interventions were effective, ethical, and socially responsible (*ibid.*).

The increasing prominence of development NGOs as agents of social change and their reliance on donor funding have led to growing pressures around their legitimacy and accountability. As they became integral to global development processes, NGOs faced heightened scrutiny about their representativeness and the alignment of their actions with the needs of beneficiaries. Questions about who NGOs represent, the impact of their actions, and the legitimacy of their influence have permeated the discourse on their role. These issues

---

[5] In this context, depolitization should not mean the dictionary definition of "removing the political character," as this process is likely impossible, but rather as adoption of the neoliberal (or Washington) consensus, dominating politics of Western countries at this time.

are exacerbated by the tension between upward accountability to donors and downward accountability to local communities (Atack, 1999; Abouassi & Trent, 2016). NGOs are often subject to "multiple accountabilities," where they must respond to different stakeholders with conflicting interests. On one hand, donors expect financial accountability, demanding transparency in how funds are used and a clear demonstration of measurable outcomes, on the other hand, NGOs are expected to maintain legitimacy by representing and being accountable to the local communities they serve. This tension sometimes resulted in what Abouassi & Trent (2016) describe as a focus on upward accountability—serving donor interests and securing funding at the expense of grassroots responsiveness (Abouassi & Trent 2016).

Some authors go as far as claiming that even though NGOs' mission statements create a normative source of legitimacy of their work, this legitimacy is then not institutionalized and organized (Ossewaarde et al. 2008). Using the example of post-Tsunami humanitarian interventions, these authors claim that as these organizations struggle to reconcile their mission with the requirements for regulatory, cognitive and output legitimacy—exemplified by the pressures from donors for transparency and quantifiable, short-term outcomes—the more the pursuit of their core objectives is obstructed (*ibid.*). Practically, this process could be illustrated by the fact that the development NGO managerial staff today spend most of their time on bureaucratic tasks—writing grant proposals or project evaluations—with minimal resources for other types of tasks.

Another source of criticism of NGO legitimacy in the face of increased donor pressures is their failure to meet their long-term goals of social transformation and poverty alleviation, even despite their substantial funding (Banks et al. 2015). Their increasing focus on short-term, measurable outcomes to satisfy donor requirements has led to criticism that NGOs are not addressing the root causes of poverty or achieving meaningful, sustainable change (*ibid.*). Attack points to instances where development NGOs were accused of failing to reach the poorest and most marginalized communities. These failures are attributed to the selective nature of their activities, driven by donor priorities, which sometimes led to a lack of focus on the most pressing needs in the regions where they operated (Atack 1999).

The perceived discrepancy between accountability to donors and to local communities and those in need further added to broader questions about the legitimacy of development NGOs. One of those is their lack of representativeness, as the internal processes of these institutions are far from democratic and the decision makers and top managerial staff, as a result of the blurring of boundaries between nonprofit, business, and government organizations, often overlap with such staff from the global 'North' (Vedder 2007). This process is broadly referred to in the literature as *mission drift*—defined as a gradual shift in an NGO's focus away from its original goals, often as a result of external pressures such as donor demands, accountability requirements, and the need to professionalize (Brass et al. 2018). This process then leads to a focus on technical solutions rather than addressing systemic issues (Ossewaarde et al. 2008). Collectively, the literature illustrates how mission drift emerges from the tension between maintaining organizational survival and staying true to the original

values and objectives that motivated the formation of NGOs (Brass et al. 2018; Ossewaarde et al. 2008; Crack 2019; Jordan & van Tuijl 2007).

The perceived loss of legitimacy towards those in need, the shift of NGO priorities further away from their original missions and closer to the needs and interests of state or IO donors, the adoption of NPA and neoliberal approaches and the blurring of lines between NGOs and the governments and other donors they serve in recent decades resulted in other hindrances to the development programs and NGO work in general. Authors like Brass et al. describe how developing country governments began to push back against foreign-funded NGOs in the mid-2000s, introducing formal and legislative restrictions on their operations and funding (Brass et al. 2018). Although significantly influenced by state relations and the nature of the authoritarian regimes in these countries, the so-called 'foreign agents laws' clearly reference the questionable representativeness and accountability of Northern-based and financed NGOs, targeting them for serving the interests of Western governments and not the mission-stated recipients of their work (Nägele 2024).

1.4. Development discourse

The concepts presented below, unified under the umbrella of *development discourse*, provide a critical lens for understanding both the explicit and implicit frameworks that guide NGOs' impact evaluation practices and, as such, provide a helpful guide in further analysis. Development discourse refers to the process of articulating knowledge and power through which specific concepts, theories, and practices for social change are formed and perpetuated (Chae 2008; Escobar 1995; Crush 1995). The importance of *discourses* in shaping social reality was inspired by the foundational works of Michael Foucault on discourse and power, identifying discourses as mechanisms of control and influence (Foucault 2009; 2019). The approach to development in terms of discourse gained most traction within the debates on Marxist dependency theory (Chae 2008) and further penetrated IR theories with constructivist authors like Wendt and Adler (Wendt 1999; Adler 2013) and post-structuralist authors asserting language and discourse of development as systematically organizing power through the subjectivity of social actors and their actions (Escobar 1995; Chae 2008; Ferguson 1994).

Within the historical development of NGOs, as previously discussed, especially the rise of development NGOs in the 1980s and the expansion of responsibilities of NGOs within the development sphere were crucial in the forming of their roles within the development discourse (Islam 2016). Building on critical approaches to development communication, the 1990s saw an increased focus on development discourse, framing it as a highly contested arena where dominant groups seek to assert control over marginalized populations (Chae 2008).

A prominent example of this process comes from Arturo Escobar. He illustrates how issues surrounding the exploitation of Colombia's Pacific rainforest were framed within the

discourse of "biodiversity," ensuring the management of genetic resources and control over genetic intellectual property. Systematically organized "facts" were being generated, largely through networks of experts linked to international agencies and northern environmental NGOs backed by G-7 countries, to offer scientific solutions aligned with predetermined objectives. He notes that this constructed knowledge deliberately sidelines the identities and cultures of Afro-Colombian and Indigenous communities in the region (Escobar 1995). This example also highlights the position of NGOs in forming, maintaining, and enacting the development discourse.

In the 2000s and 2010s, the role of NGOs was central in the international aid regime, and, as Islam notes, as their influence grew, concerns over accountability and dependency on foreign donors intensified, with critiques focusing on upward accountability to donors rather than downward accountability to communities (Islam 2016; Murtaza, 2012). Escobar's research into the mechanism of the development discourse process in Colombia's Pacific rainforest exploitation serves as an example of this upward accountability (as discussed in the previous chapter) within the development sphere.

A more recent account from the post-structuralist perspective about the impact of dominant development discourse on local communities is illustrated in Tania Murray Li's *Will to Improve*. Critically examining development discourse through Foucault's concept of governmentality, she examines how the drive to "improve" becomes a mechanism of control within development practices (Li 2007). Criticizing development practices presented as neutral, technocratic solutions, even though they are deeply embedded in broader power structures, she describes the role of expertise, where external actors—development agencies or NGOs—define problems and prescribe solutions that align with donor agendas and global priorities, sidelining local knowledge and autonomy. Through exclusionary practices and narratives, she shows how the development process is expropriated and carefully managed to maintain control over the development process while often neglecting structural causes of inequality and undermining local livelihoods and agency (*ibid.*).

Along with the dominance of donor accountability highlighted earlier, the dominant development discourse within the international development NGO sector today is exemplified by several other factors. Authors like Eyben highlight the growing emphasis on *evidence-based, results-driven development* and its implications for how NGOs approach impact evaluation, highlighting the limits this places on long-term, transformative change (Eyben 2013). As discussed in the previous chapter, the impact of market-oriented solutions also plays a critical role in today's development discourse, as highlighted by Roy and others (Roy 2010).

Authors like Cornwall and Brock analyze how buzzwords like "participation," "empowerment," and "poverty reduction" are used in development discourse, often shaping practices that may not always align with meaningful, grassroots empowerment (Cornwall and Brock 2005). Concepts set by dominant multilateral bodies like the UN or OECD can also serve a similar function. In recent years, for instance, these were embodied by the

Millennium Development Goals and are now exemplified by the Sustainable Development Goals (SDGs) (UN 2015). These then serve as a guiding framework for many international NGOs (Sachs 2012).

Practically, the notion of development discourse highlights that impact evaluation practices are not neutral or purely technical but are shaped by specific discursive frameworks. It encourages looking at how NGOs' practices and metrics for evaluating impact may be influenced by dominant narratives about development. For instance, discourses emphasizing measurable, quantifiable outcomes might favor evaluation methods that prioritize short-term, tangible impacts over long-term structural change. Development discourse also directs attention to the kinds of language used in NGOs' mission statements, strategies, and public documents, as this language often reflects broader assumptions about what constitutes "successful" development. By examining this language, the framing of NGOs' goals can be better understood in ways that align with or challenge mainstream development ideals. This discursive analysis informs how evaluation practices are interpreted in terms of what they prioritize and what they may overlook.

In analyzing and exploring background knowledge, the key concepts of development discourse help in addressing the implicit assumptions, values, and "taken-for-granted" ideas that shape NGO practices. Reminding also to focus on how these implicit beliefs are culturally and historically constructed, often embedded within the expectations of donors, international agencies, and the broader development industry. When discussing the resulting implicit practices and background knowledge uncovered within this analysis, a potential dissonance between what NGOs say and what they do can be interpreted with the use of this concept. This process would also highlight the process of formulating the development discourse from implicit practices into the process of impact evaluation.

1.5. Impact evaluation as practice

Impact evaluation practice presents a key concept for the scope of this research, but as such, in line with the theoretical overview presented above, it is a multifaceted and varied practice. The OECD Development Assistance Committee (DAC) defines impact evaluation as the systematic assessment of the long-term changes—positive or negative, intended or unintended, and primary or secondary—that result from a development intervention (OECD 2023). However, authors such as Belcher and Palenberg point to the fact that there is a widespread inconsistency and overlap in definitions of key terms such as "outcome" and "impact" in development evaluation across major definitions used by entities like the OECD, UN agencies, and development banks, creating challenges for mutual understanding and application (Belcher & Palenberg 2018). This lack of definitional rigor not only complicates evaluation practices but also shapes divergent understandings among development NGOs of what "impact" means and how it should be assessed within the constraints of the aid system (*ibid.*).

For instance, while the OECD defines impact as long-term changes—both intended and unintended—attributable to interventions, this definition often adopts an "intervention perspective" that overemphasizes linear causality and full attribution. This perspective is frequently at odds with the complex, systemic realities of development contexts where multiple factors interact. Moreover, the ambiguity in temporal qualifiers like "short-term" and "long-term," as well as the inconsistent differentiation between outcomes and impacts, undermines evaluative consistency and accountability (*ibid.*).

According to one of the most prominent handbooks on impact evaluation by the World Bank, "impact" is defined as the long-term, ultimate changes that result from an intervention, higher-level goals, or objectives the program seeks to achieve, often beyond the immediate control of the project implementers. For example, reduced poverty or improved health at a national level. Impacts often require systemic changes and are influenced by multiple factors beyond the specific program being evaluated. (Glewwe & Todd 2022). This document defines the practice of impact evaluation as the systematic assessment of the causal impact of a project, program, or policy on outcomes of interest. This practice aims to establish whether interventions achieve their intended objectives by determining the relationship between the intervention and observed changes (*ibid*.).

To aid the process of impact evaluation in practice, a negative delamination of impact can be helpful for the process of analysis. Firstly, "impact" is distinct from "outputs" (e.g., number of workshops conducted or policies enacted) and even "outcomes" (e.g., changes in skills or behaviors of participants); according to the research cited, these are the most common confusions (Belcher & Palenberg 2018; Chianca 2008). Impact is also not linear; it cannot be understood as causal chains directly linking interventions to changes. Impact also does not include only intended changes; and should include both positive and negative unintended consequences. Impact also cannot be fully captured by short-term assessments, missing its long-term nature. Impact also cannot be defined by a single study, and replication is necessary to untangle the whole picture (Belcher & Palenberg 2018; Chianca 2008). Impact cannot be assessed independently of the sociopolitical, cultural, or economic systems in which interventions occur (Belcher & Palenberg 2018; Glewwe & Todd 2022).

One of the most widely adopted, standardized frameworks for assessing development interventions is the OECD DAC Criteria—relevance, effectiveness, efficiency, impact, and sustainability, expanded in 2019 with coherence (OECD 2019). As Chianca highlights, while the DAC criteria provide an internationally recognized foundation for evaluations and are particularly influential for professionalized development NGOs, their application often reflects significant shortcomings that affect both the conceptual and practical dimensions of impact evaluation (Chianca 2008). Chianca underscores the inherent tension in the criteria's application, particularly in their donor-centric orientation. For instance, relevance, a criterion intended to assess alignment with beneficiaries' needs, often prioritizes the objectives of donors or governments, thus marginalizing the perspectives of the communities targeted by interventions. Most importantly, in line with findings by Belcher and Palenber, the conceptualization of impact, often conflated with outcomes, poses additional challenges.

Chianca critiques the prevailing emphasis on linear causality and full attribution, which overlooks the interconnected nature of development contexts (*ibid.*).

In line with increased pressures on NGO accountability and questioning of the mantra of NGO effectiveness, the amount of impact evaluations has grown in recent decades. White notes the growth from less than 20 published studies a year before 2004 to 120 a year during 2012 (White 2014). However, he highlights that compared to the number of programs, countries, and sectors addressed, this is still a very low number with insufficient replication. It is crucial to note that these are predominantly randomized controlled trials (RCTs) often conducted by external evaluators[6]. While these are the preferred, rigorous method, they are not the norm in impact evaluation in international development NGOs in general, as White notes; often RCTs are not feasible for all interventions due to logistical and political constraints (*ibid.*). Oftentimes, they can be cost-prohibitive or unreasonable compared the scope of the particular initiative. Lastly, there is a noticeable lack of replication, and when replication is conducted (like in the case of impact assessments of deworming on school attendance), it can yield contradictory results (Miguel & Kremer 2004; Davey et al. 2015).

Hands-on impact evaluations are done by measuring the difference between the treatment group and the control group, aiming to control for other variables (Gertler et al. 2016). Cook and Campbell (1979) state that conducting an impact evaluation requires meeting three key conditions—covariation, temporal precedence, and the elimination of plausible alternatives—before a cause-effect relationship in a development intervention can be established. Conventional methods used for impact evaluations are interrupted time-series analysis, difference-in-difference analysis, regression discontinuity analysis, and cross-section analysis (Rahman & Farin 2019).

Both White and Glewwe & Todd emphasize the consideration of *counterfactuals* in impact evaluation, asking what would have occurred without the intervention (Glewwe & Todd 2022; White 2014). Using hypothetical scenarios of what would have occurred in the absence of the intervention, such impact evaluation should serve as a benchmark against which the effects of the intervention can be measured (Glewwe & Todd 2022). As mentioned, among the methodologies for constructing counterfactuals, Randomized Controlled Trials (RCTs) are widely regarded as the gold standard due to their ability to minimize bias through random assignment, ensuring comparability between treatment and control groups. Alternative methods such as difference-in-differences, propensity score matching, and regression discontinuity designs are frequently employed, each offering distinct approaches to approximate the counterfactual in diverse contexts (*ibid.*).

Rahman and Farin critically assess the current practices in impact evaluation, first pointing to the attribution problem (Rahman & Farin 2019). When dealing with outcomes and especially impact, many more variables come into play, making attribution of potential

---

[6] Organisations such as International Initiative for Impact Evaluation (3ie), university research bodies such as the Trinity Impact Evaluation Unit (TIME) or organizations and donor bodies such as CIDA or USAID.

results to program effects difficult. They are using the *micro-macro paradox* (first mentioned by Mosley (1986) and later confirmed by Boone (1996) to illustrate. These studies uncovered that even if the vast majority of micro-assessment reports alluded to resounding success in meeting the desired impacts, subsequent macro-level econometric studies on the combined effects of all interventions conducted at the country level often showed disappointing results. Simply put, the overall impact of a development program is smaller than the combined effects of its individual components. This could be the effect of the *aid curse*, as negative economic effects only become evident on a macro level (such as inflation, Dutch disease, or corruption) (Rahman & Farin 2019).

Practical impact evaluations of development NGOs are closely linked to the concepts of the theory of change, also called result-based management, a management approach focused on achieving development results through effective planning, implementation, monitoring, learning, and reporting, which often refers to the changes that occur in a society or among beneficiaries as a result of an intervention (*ibid.*). This logical model is beneficial since, as Rahman and Farin note, interventions often fail simply due to a poorly designed logical framework. But in reality, even if the logic of intervention is spelled out, assessments of interventions tend to be different depending on analysts and evaluators who operate with different frameworks and models (Rahman & Farin 2019; White, 2005). The authors further add that, as their literature review reveals, evaluations can differ widely based on their purpose, the questions posed, the target audience, the conceptual frameworks utilized, the intervention's context, the type of analysis conducted, and the specific methods and approaches chosen by the evaluator (*ibid.*).

1.6. Practice theory framework

Practice theory in IR provides a key framework for the objectives of this research, building on the work of Adler and Pouliot with key practical refinements by Bueger (Adler & Pouliot 2011; Adler-Nissen & Pouliot 2014; Bueger 2014). This approach turns its focus on practices—defined as *competent performances*—socially meaningful patterns of action that embody and solidify *background knowledge* and *discourse* on the material world, allowing for structures to be stable or evolve and agents to reproduce or transform structures. With this approach, the authors are turning their focus on what practitioners do in their everyday actions and how these practices collectively shape world politics. This approach aims to offer a pluralistic framework, bridging divides between existing IR theories and fostering interparadigmatic dialogue. Moreover, its advantages lie in conceptualizing short-term societal change (Neumann 2002), getting closer to the everyday activities of those speaking, writing, and doing politics, and developing analytical approaches resonating with other than scholarly communities (Bueger 2014).

Practice theory, which positions practices as the smallest unit of analysis, has gained prominence across disciplines, originating with Sherry Ortner's (1984) anthropological work

and Pierre Bourdieu's theorizing. It has since expanded to fields such as organization studies (Nicolini et al. 2003; Golsorkhi et al. 2010), history (Spiegel 2005), and political science subfields, including policy studies (Hajer & Wagenaar 2003) and international relations (Neumann 2002; Adler & Pouliot 2011).

practices have long been a focus within IR, evident in Kissinger's work on diplomacy and in contributions from the English school, rationalists like Thomas Schelling, and liberals such as Keohane and Nye, even before the 1980s (Adler & Pouliot 2011). However, practice theory, as articulated by Adler and Pouliot, draws on post-structuralism, particularly Foucault's emphasis on everyday practices, and builds on constructivist and Bourdieusian insights into practical reasoning and deeds. Unlike earlier frameworks where practices were secondary to other concerns, practice theory foregrounds them as the central analytical unit. Neumann's linguistic turn (Neumann 2002) further reinvigorated the focus on practices by integrating language and meaning into the IR framework, bridging textual and practical dimensions into a unified theory.

Practice theory differentiates behavior, action, and practice to capture the progression from material deeds to socially meaningful, patterned activities. Behavior refers to actions devoid of meaning, while action introduces intent and significance within social contexts. Practices, however, extend further by being repeated, socially meaningful performances embedded in structured contexts, sustained by shared background knowledge. For example, running as a form of physical training reflects a socially organized practice, distinct from aimless running (behavior) or chasing a thief (action).

Practices are characterized by regularity over time and space, dependence on *background knowledge,* and their capacity to bridge material and discursive realms. This makes them dynamic: practices simultaneously stabilize existing structures and enable transformation. They evolve through a lifecycle of generation, diffusion, institutionalization, and decline, responding to changing social and material conditions. Practices also function as both *explanandum*—phenomena to be explained, such as the emergence of multilateral diplomacy—and *explanans*, driving broader social transformations by reshaping norms, power dynamics, and knowledge systems. This concept relates to Neumann's *discursive performances*—emphasizing that practices are inherently discursive, as they embody, enact, and sustain discourse through action (Neumann 2002).

Practice theory operates with the key concept of *background knowledge,* which is crucial for the aim of this research. Background knowledge is a foundational element of practices, encompassing the often-tacit understandings, skills, and dispositions that enable competent performances. Unlike explicit knowledge, background knowledge is deeply practical, oriented toward action, and resembles skill rather than formalized rules or norms. Bueger lists various conceptions of background knowledge, signifying its multifaceted nature, as:

- understanding, know-how, emotional states, and motivational knowledge (Reckwitz 2002),

- practical understandings (know-how), rules (explicit instructions or prescriptions for action), and teleoaffective structures (arrays of goals, projects, emotions, and acceptable uses of things within a practice). (Schatzki 2005),
- *habitus* (implicit knowledge embodied in individuals) and *doxa* (implicit, field-specific knowledge within social formations) (Bourdieu 1977), or
- implicit knowledge organized as narratives that provide stability to practices over time (Rouse 1996).

Background knowledge is both embodied and enacted within practices, shaping and being shaped by their iterative performance. This knowledge underpins the shared expectations and dispositions that make social interaction possible, providing the interpretive framework for what is deemed competent or meaningful. Importantly, background knowledge is not static; it evolves through processes of learning, contestation, and adaptation, often becoming visible and subject to debate during moments of crisis or transformation. By integrating material and discursive dimensions, background knowledge binds structure and agency, functioning as the invisible scaffold that sustains and reproduces social practices over time.

Another important concept in practice theory is *communities of practice*—social collectives where practices are developed, maintained, and transformed. These communities consist of individuals or organizations sharing a domain of knowledge, mutual engagement, and a repertoire of communal resources such as norms, routines, and discourses. Within these communities, practices are socially structured and transmitted, enabling members to interpret and perform them competently. Background knowledge then aligns members' actions and fosters a shared understanding within these communities. Practices within these communities evolve through iteration, contestation, and adaptation. The authors illustrate diplomatic networks as an example of such a community, but in our case, communities of NGO officers and project managers boast similar characteristics.

Practice theory moves the focus from evaluating outcomes or end results to examining *how* practices, such as impact evaluation, are performed. This research uses practices as *explanans*, driving broader social transformations by reshaping norms, power dynamics, and knowledge systems. Specifically, it aims to look at practices of impact evaluation and the background knowledge guiding these practices to uncover the hidden meanings in assessing impact that create, manifest, and reproduce development discourse. Emphasizing the dual role of practices as shaped by structural constraints (e.g., donor requirements, professional norms) and as agents of potential change directs the research to analyze not just what NGOs do but how these actions reflect, reinforce, or challenge the broader aid system.

In line with practice theory's focus on the interplay between articulated and tacit knowledge, RQ1 begins with an empirical mapping of observable practices, which informs RQ2's deeper exploration of the meanings and assumptions that underlie them, creating a dynamic process of inquiry. RQ1 centers on uncovering the explicit, observable elements of these practices— methods, routines, and decision-making processes—while recognizing that these are not merely technical acts but are infused with organizational norms, values, and broader

systemic influences. By conceptualizing practices as carriers of meaning, practice theory prompts RQ2 to delve beyond the surface-level explicit knowledge (mission statements, annual reports) to uncover the implicit *background knowledge* that sustains and informs those practices. This includes taken-for-granted assumptions about what constitutes "impact," how it is measured, and why certain evaluation methods are preferred over others.

2. Methodology
2.1. Introduction

This chapter outlines the methodological approach used to explore how development NGOs conceptualize and measure impact within the aid system. Building on the theoretical background presented in earlier chapters, two key research questions guide this inquiry: first, how professionalized NGOs evaluate the impact of their work (RQ1), and second, what background knowledge underlies these evaluation practices (RQ2). The research questions reflect practice theory's emphasis on uncovering both explicit and implicit meanings embedded in everyday routines. Accordingly, the study focuses not only on how NGOs articulate impact through public documents and donor communication but also on the tacit assumptions that inform these practices. By drawing on short-term project models to capture long-term goals, the chapter allows for later exploration and nuanced analysis of the processes that shape NGOs' understanding and assessment of impact.

To operationalize these inquiries, the chapter outlines directed qualitative content analysis (DCA)—used to answer RQ1—alongside a praxiographic lens building up on these findings to answer RQ2. DCA provides a structured yet flexible framework for analyzing textual materials (mission statements, annual reports, and strategic documents) while remaining anchored to established literature. This method systematically identifies key concepts drawn from the theoretical framework and refines them based on the collected data. Meanwhile, praxiography, influenced by Bueger's methodological refinements of the theory outlined by Adler and Pouliot, offers a complementary approach to studying the implicit knowledge that underpins day-to-day practices.

Next, this chapter discusses case selection and concludes with reflexivity and limitations to provide a transparent account of the study's scope. Three European-based development NGOs—People in Need, Helvetas, and Concern Worldwide—were purposefully chosen due to their shared membership in Alliance2015, comparable budget sizes, and operational overlaps. While this homogeneity allows for a deeper analysis of a distinct community of practice, it also narrows the applicability of findings. Acknowledging the interpretive nature of practice theory, the chapter highlights the reflexive stance, noting how reliance on expert interviews and document analysis may privilege consciously articulated knowledge over unspoken norms. Limitations also stem from the absence of participant observation, potentially leaving some practices out of view. Nonetheless, by balancing directed content

analysis with praxiography, the study aspires to shed light on both the explicit and implicit dimensions of impact evaluation.

2.2. Research objectives

The objective of this thesis is to analyze how development NGOs that strive for high-level impacts, but operate within the constraints of the aid system, as discussed in the theoretical background, understand and evaluate the impact of their work. Following practice theory, the research focuses on impact evaluation as a standardized practice, one which inherently includes in it the meanings not only about how impact is best assessed but also about what impact is in the first place.

The research questions are the following:

(RQ1) How do professionalized development NGOs evaluate the impact of their work?

The objective of this research question is to provide an account of how NGOs operating primarily in *short-term* project models assess the *long-term* objectives of their development projects or programmes on the level of concrete impact evaluation practices. This includes objectives both on the organizational level as well as objectives of specific interventions. Within this question, specific routines, methods, and decision-making processes involved in impact evaluation are investigated. The goal is to uncover the explicit, articulated knowledge of NGOs about impact and impact evaluation, which should be evident in mission and vision statements, annual reports, strategies, project proposals and other documentation directed at the public and/or the donors.

(RQ2) What background knowledge underlies the impact evaluation practices in professionalized development NGOs?

Following the conceptualization of practices in practice theory as carriers of meaning, the objective of this research question is to examine the background knowledge underlying the impact evaluation practices identified in RQ1 – the implicit, taken-for-granted understandings about what impact is and how it is best assessed. This question will take as its starting point the explicit, articulated knowledge of NGOs about impact and impact evaluation (evident in mission and vision statements, annual reports, strategies, project proposals and other documentation directed at the public and/or the donors). Then, the background knowledge will be reconstructed by an analysis of consistencies and inconsistencies between what is expressly said, what is done (practices identified in RQ1) and what is reflected on in submitted internal documentation and during in-depth interviews.

2.3. Operationalization and analytical framework

2.3.1. Directed Qualitative Content Analysis (DCA)

A structured qualitative approach was selected as a tool to review materials and aid in answering RQ1. Aiming to stay close to data sources but utilise the theoretical framework overviewed in the previous chapters led to selecting a structured approach, as it allows for the necessary depth and flexibility in conforming to the specific data analysed while aiming to stay close to existing theoretical frameworks and mitigating potential biases or diversion that poses a threat in unstructured, impressionistic or intuitive qualitative approaches (Hsieh & Shannon 2005; Schreier 2012; Neuendorf 2017).

The literature offers three distinct approaches to content analysis: conventional, directed and summative (Hsieh & Shannon 2005). Conventional content analysis was not used as it is appropriate when existing theory or research literature on a phenomenon is limited (*ibid.*), which is not the case with *impact evaluation* for development NGOs. Summative analysis, on the other hand, focuses on quantifying the uses of certain words in text, followed by content analysis, blending qualitative and quantitative approaches. These analyses are frequently employed by researchers focussing on specific manuscript types in a particular journal or specific content in textbooks (*ibid.*). For our specific case, an approach with this level was not suitable, especially as RQ1 serves as an overview and a stepping stone to answering RQ2, which is the real focus of this research.

The directed qualitative content analysis (DCA) lies in the sweet spot between conventional and summative analyses. Its goal is to validate or conceptually expand a theoretical framework or theory. Existing theory or prior research can help refine the research question, offering predictions about the variables of interest or their relationships. This, in turn, aids in developing the initial coding scheme or identifying connections between codes (Hsieh & Shannon 2005; Mayring 2000).

The process of DCA begins by identifying key concepts or variables as initial coding categories, followed by determining definitions for each category with the help of existing theory (Potter & Levine-Donnerstein 1999). After establishing key predetermined codes, the coding of texts can immediately begin, with data that cannot be initially coded set aside for further analysis to determine whether they represent a new category or a subcategory of an existing code (Hsieh & Shannon 2005). What is then suggested is the use of rank order comparisons of the frequency of codes, followed by using the theory to guide the discussion of findings (*ibid.*).

2.3.2. Praxiography: the practical application of practice theory

Although described in great detail in the foundational book by Adler and Pouliot (2011), practice theory has been driven mainly by abstract, scholarly concerns motivated by

epistemological and ontological contemplation, with sparse focus on practical methods (Bueger 2014). In reaction to this vacuum, Christain Bueger proposes a methodology dubbed 'praxiography' that suggests forms of analysis produced by practice researchers, along with practical guidelines on how to carry out praxiographic research (Bueger 2014; Mol 2003). He offers a repertoire of methodological approaches to researching practices, informing the process of answering RQ2 in uncovering the *background knowledge* underlying *impact evaluation practices*.

Bueger addresses several problems of praxiographic research. The first problem is its focus on implicit (background) knowledge which is rarely verbalised and not easily readable, as it is evident only in the carrying out of practices. He deconstructs practices as requiring (1) forms of bodily and mental activities, (2) artefacts, and (3) background knowledge giving practices their meaning. What gives practices their materiality are *artifacts* that are necessary to carry out such practices (such as pen and paper giving materiality to the practice of writing). Overall, practice theory operates in two dimensions–observation and interpretation- but of the three elements of practice, only forms of activities and artefacts can be observed, background knowledge can only be interpreted. A methodological approach will thus involve analysing the observable aspects of practices and their interplay to reconstruct background knowledge.

The process of reconstruction of this knowledge is reconstructing meaning, hence, an interpretative and qualitative process, involving observation followed by interpretation. To reconstruct implicit knowledge will require considering articulated meanings, utterances, actions or the handling of objects and artefacts. To avoid forcing meaning upon the studied practices, praxiography attempts to identify moments in which participants in practice tend to articulate implicit meaning themselves.

Although praxiography does not explicitly focus on articulated meaning, Beuger notes that much of the research *de facto* draws on primary data (such as explicit rules, classification, cultural codes, metaphors or discourse. Analysing this data, as Bueger notes, then brings praxiography to a close relationship to discourse analysis, but praxiography will primarily attempt to gather first-hand observations of speech and only consider documents when sayings and doings cannot be observed directly (Bueger 2014).

The design of this study pragmatically reflects this reality by first gathering and analysing explicit impact evaluation materials in RQ1, followed by a more purposeful praxiographic approach in RQ2 based on gathering first-hand observations of speech via interviews. In contrast to discourse analysis, however, primary data will be considered in relation to RQ2 as documents that relate to bodily material practice outside of the text.

The second problem of praxiography that Beuger identifies is the problem of scale and formulating statements that are relevant beyond the immediate local contexts. Practice theory often aims to understand large, multinational *communities of practice* through very small acts of the everyday practice of a few of its members. Three answers can bridge this obstacle: (1) the study of practice is an open scale, with smaller practices nested in larger

ones and interacting regularly, thus it is relevant to focus on any practice; (2) the study of practice does not need to entail the full complexity of practice, often an overview will be sufficient; and (3) praxiographes often reject the idea of natural scales (micro, macro, local, global…), aiming to transcend these, claiming that practices can be both local and global at the same time.

The third problem is the problem of contingency and unruliness of practices compared to more stable systems governed by logic or laws. Practices are repetitive but also constantly changing and evolving. This problem can only be addressed by reflexivity of the findings of praxiographic studies. Additionally, researchers distinguish between minor adjustments and major ruptures in practice (Shatski 2002).

Beuger then proposes three strategies of praxiography (1) sites, (2) crisis and controversies, and (3) following objects.

For *sites*, a special focus is given to the structure-making sites, which, although local, transcendent their local setting in their macro-structural impact on their corresponding systems. Although a focus on sites of impact evaluation–such as NGO headquarters–could have been an option, access and field research is beyond the scope of this research. Perhaps more importantly, much of the usual impact evaluation practice in professionalised development NGOs nowadays takes place online, not making the site approach suitable. Nonetheless, site relevance might still emerge from studying speech and artifacts.

Crises and controversies have long been favoured by constructivists for their insights on understanding knowledge and the introduction of new ideas. Beuger notes that in such instances, implicit knowledge often becomes explicit, in argument about new re-structuring of systems, and as this re-structuring occurs, new insights on old systems might emerge. Similarly, however, observing crises was not suitable for the scope and abilities of this research.

Therefore, it results mainly in the third strategy—following objects—meaning objects and technologies but also language artefacts such as concepts or metaphors, as they can be crucial containers of practice. The process begins by a detailed study of an artefact, followed by an interpretation of which practices are inscribed in it. This approach is dubbed objective ethnography. In the use of technology, Burger, among others, mentions an example of how using information technologies spurs the practice of public diplomacy. Documents are then described as tools of connection of sites with the world, the core materials of policy-making and essential in the life of organisations. Therefore, they are the fundamental material of political practice and offer valuable insight. Lastly, concepts, metaphors, or other linguistic constructs are signifiers that reveal practices.

2.4. Data collection methods

Praxiography as proposed by Bueger draws on three primary methods of data collection–participant observation, document analysis and expert interviews (Bueger 2014). Since participant observation is beyond the scope of this research, it focuses on the latter two.

A set of *expert interviews* from each of the observed organisations was conducted to unravel the implicit structures of meaning and the reasons why they are widely used. Burger states it is important to clarify the relationship between the interviewee and the practice. In our case, all interviewed experts were either conducting impact evaluations themselves, in the position of advisors for monitoring and evaluation or overseeing the process of monitoring and evaluation in these organisations. All of these fall into the two necessary classifications for praxiography interviews–individuals who either participate in the practice on an everyday basis or spend considerable time observing the practice.

Interviews were conducted according to the *interview to the double* practice suggested by Nicolini (2009). In such an interview, the researcher and the interviewee reconstruct implicit meaning together and coproduce the interpretation of practices. The interviews with senior advisers also helped the interpretation process, as suggested by Bueger (2014), but the lines between the two are often not clear.

The structuring of the interview was informed by the types of data produced: (second-hand) interpretations of implicit knowledge based on observations of the interviewee.

As for *documents*, Burger identifies three types–(1) handbooks and manuals, (2) documents describing practices, and (3) visual artefacts. The first two were analysed in this instance. Manuals describing how to conduct *impact evaluation* were collected from all of the analysed cases, paying close attention to the fact, as Bueger notes, that manuals are not the same as practices themselves, but can give us important clues about the practices and the knowledge that informs them. A reflexive stance towards the idealised nature of these documents was taken and this nature was contrasted with the practical reality emerging from interviews. The second type of document was mainly represented in the form of activity reports on impact evaluation, proceeding with the same reflexive analytical approach.

2.5. Case selection

The selection of cases in this research reflects a deliberate and purposeful approach, guided by the theoretical framework of practice theory and the study's objectives. Practice theory emphasizes the analysis of practices as carriers of socially meaningful patterns of action, shaped by background knowledge and context (Adler & Pouliot, 2011; Reckwitz, 2002). As such, the choice of cases must facilitate an in-depth examination of the standardized practices and underlying meanings within impact evaluation processes. To this end, three European-based development NGOs—People in Need, Helvetas, and Concern Worldwide—were selected for their shared characteristics as professionalized, Western-based organizations and their institutional contexts and practices.

The selection of cases in qualitative research, as emphasized by Yin (2018) and Patton (2014), must align with the research objectives and theoretical framework. In this study, the following criteria were employed: (1) Relevance to Research Objectives, (2) Diversity and Variation, (3) Comparability and Theoretical Replication, and (4) Practical Feasibility.

The selected cases are relevant to the research objectives since they are all NGOs– professionalized development organizations with established practices for evaluating the impact of their work. These organizations operate within the constraints of the aid system, engaging in short-term project models while striving for long-term development goals. This alignment ensures that the cases provide insights into the explicit and implicit dimensions of impact evaluation practices, addressing both RQ1 and RQ2

The cases were chosen to mitigate variation in organizational size, network affiliation, operational contexts and base location. All of the organizations (People in Need, Helvetas, and Concern Worldwide) are ranked as middle-sized development NGOs (with an operating budget between USD 100 million and 1 billion), are based in Europe with operations in Africa, Asia, Middle East and Eastern Europe, with many overlaps. Moreover, all are members of Alliance2015, a strategic network fostering collaboration among European development organizations.

This case selection provides benefits and drawbacks. For their many similarities, a study designed this way allows for a deeper dive into the *impact reporting* practices within a more narrow *community of practice* and, with varying cases from the same sub-category within development NGOs, practices within this community can be described in more detail. On the other hand, due to the homogeneity of cases, the results of this study will only be applicable to this sub-group, with limited opportunity for generalisation.

2.6. Reflexivity and limitations

Most broadly, using a qualitative approach includes inherent limitations such as smaller sample sizes, which can restrict the generalizability of the results compared to quantitative methods. The subjective nature of data collection, driven mainly by abstract theory, and analysis may introduce researcher bias, and the lack of standardized metrics can make replicability and comparison across studies challenging. Moreover, as qualitative methods often require extensive time and resources for in-depth analysis, this research was limited to a smaller number of cases.

Using theory within DCA has some inherent limitations in that researchers approach the data with an informed but, nonetheless, strong bias. Hence, researchers might be more likely to find evidence that is supportive of the theory rather than nonsupportive (Hsieh & Shannon 2005). Also, overemphasis on the theory can blind researchers to contextual aspects of the phenomenon (*ibid.*). Subsequent critical examination of results with this knowledge can mitigate some of these limitations.

A set of this study's limitations stems from its heavy reliance on practice theory, carrying on its general limitations. Adler and Pouliot highlight that defining what clearly constitutes "practice" is quite difficult due to its broad applicability across disciplines. This can lead to vagueness in its application to accommodate diverse perspectives (Adler & Pouliot 2011). Even though significant work has been accomplished in his text, Bueger still criticises the lack of clear methodological guidelines for conducting research within a practice-theoretical framework (Bueger 2014). This type of research approach is still vaguely defined and the lack of methodological rigour allows for vastly diverse interpretations and approaches, therefore, a praxiographic study still relies heavily on the researcher. As this is our first practical experience with conducting practice theory, this study is running the risk associated with a relatively loose methodology. The interpretive nature of praxiography also means that the researcher's biases and perspectives may shape the reconstruction of implicit knowledge. Even when reflexivity is included, there may be limits to mitigating these influences.

Adler and Pouliot (2011) and Schatzki (2002) emphasise the limits of practice theory's ability to address macro-level structures or cross-context comparisons due to its focus on context-specific practices. Schatzki adds that practices are contingent and fluid, resisting fixed categorization which makes it difficult to develop consistent explanations. As noted in the explanation of praxiographic approach, an evident weak point lies in the interpretation of resources to reconstruct background knowledge (Bueger 2014). Such interpretation can miss the mark, especially when key context is missing. And the role of expert interviews as primary data sources as well as aiding and coproducing the interpretation of practices could easily skew this process towards their particular biases.

Adler and Pouliot further add that the micro-focus of practice theory risks overlooking larger structural and systemic dynamics, along with the difficulty of conceptualizing and measuring social and political change within the practice framework (Adler & Pouliot 2011; Schatzki 2002). Bueger highlights a potential undervaluing of explicit rules, norms, and articulated knowledge, as by privileging tacit practices, it may fail to account for the role of formal institutions and codified systems in shaping behaviour (Bueger 2014).

The selection of studied cases also introduces limitations. Their temporal and characteristical similarities limit the applicability of the results of this study on the particular sub-group of development NGOs' impact evaluation at this particular time. All selected NGOs belong to Alliance2015, which fosters collaboration and shared practices. This affiliation might lead to overemphasis on commonalities at the expense of uncovering diverse or divergent practices within the sector.

While a critical lens for understanding the impact and NGO development work has been outlined, the focus of this study on Western NGOs and the selection of experts might introduce a Western-centric bias. Due to the limitations in the scope and resources for this study, other methodological approaches to praxiography–like the study of audiovisual materials, and especially participant observation, which is the most widely used approach– could not be performed. Without participant observation, much of the implicit, embodied,

and context-specific knowledge about practices may remain inaccessible, relying heavily on interviewee interpretations and self-reported accounts. Relying mainly on expert interviews makes the participants' ability and willingness to articulate implicit knowledge a key point. This could result in overemphasis on what is consciously verbalized, potentially missing deeper understandings.

3. Analytical part

3.1. Overview of the Results

3.1.1. Types of Evaluations

The purpose of this section is to introduce the general approaches to monitoring and evaluation (not only for measuring impact), outlining key topic clusters that will be explored later in this chapter. In the cases examined, organizations assess their overall, long-term impact alongside other results (e.g., organizational efficiency, outputs, and outcomes) using a variety of evaluation practices. While the approaches employed across these organizations share significant similarities—the principal being their dependence on institutional donor funding that is associated with specific evaluation requirements—their specific organizational cultures and internal structures also result in notable differences. Elements of institutional learning and best industry-specific practices also contribute to similarities.

A central feature of all cases is the end-of-project evaluation, typically guided by the OECD DAC criteria, sometimes used alone in other times accompanied by other methodologies[7]. Mid-term evaluations, usually required for larger projects, follow less strict methods, often adapting internal reporting standards or employing modified DAC criteria[8]. Although intended to serve both donor compliance and organizational learning, the former frequently takes precedence. This can be illustrated by the fact that when end-of-project or mid-term evaluations are not required (typically for smaller projects falling under a certain budget or timeframe threshold, although these are the minority), the NGOs choose to perform different evaluation methods in line with their preferences, like institutional learning or more practical *results-based management* approaches favoring quality improvement over impact assessment.

The aggregation of reporting results for further use is influenced by organizational structure. Some follow a centralized approach in which monitoring and evaluation units at headquarters play a major role, while others favor a decentralized model, granting country or project teams greater autonomy.

To gather findings and distill lessons learned, organizations use a variety of tools. Most organizations require that evaluation findings be stored on shared platforms, thereby

---

[7] Approaches like beneficiary assessment, outcome harvesting, or capitalization were mentioned.
[8] Notably, the impact category would be absent.

facilitating broader institutional learning and review by PMs, managers, and other staff. This step is followed by their inclusion into further project planning, strategic planning, or other internal processes. Some also engage in the creation of so-called "meta-studies" or "flagship summary documents," reviewing the results of multiple projects and distilling the main lessons learned. There are also cases of the results of several related and/or similar in-scope projects evaluated together in one study, with an aim to optimize resources and generate strategic learning[9]. In addition to these, some organizations engage in annual or periodic reviews of their programs too.

Concerning impact evaluation, it is crucial to note that none of these studies allow for a methodologically sound assessment of impact as defined in the OECD DAC. Studies in this format usually evaluate at the project's end or during its implementation, when *long-term* impacts cannot yet be observable. Additionally, resources are typically minimal, far less than what would be necessary to conduct a true counterfactual study, an RCT study, or hiring expert external evaluators.

However, there are exceptions. Some organizations do engage in *ex-post* impact evaluations that employ a more sophisticated methodology and are done by expert external evaluators. Several justifications guide the utilization of these *proper impact assessments*: (1) They are employed for flagship projects that are core to the organizations' operation or are of special importance, (2) some donors occasionally require ex-post impact evaluations after passing a specific threshold[10]. But in practice, these studies are done extremely sparsely, for less than one percent of projects. In practice, this means one rigorous, *ex-post* impact evaluation for every couple of years or several hundred projects, mainly due to their *limited applicability* or *lack of resources.*

3.1.2. Who conducts evaluations

An important distinction between the evaluation performers allows for finer distinction and uncovers important background knowledge guiding how the subjects engage in impact evaluation.

In practice, different evaluation studies are conducted by:

1. internal staff familiar with the project or program—i.e. project managers, program leads, or other executive staff

2. internal staff specialized in monitoring and evaluations—monitoring officers or advisers, or

---

[9] As intervewee from Case 1 mentions, this is also used for projects that "slip through the cracks" or there they "maybe don't end up having budget for them".

[10] USAID was mentioned as an example of a donor that requires such external evaluation every few years for projects it funded.

3. external monitoring specialists.

Practically, however, the blurring of the lines between these occurs quite regularly, especially due to the expert knowledge necessary to evaluate specialized projects. In reality, internal monitoring officers are often former project managers who were implementing these projects earlier. Similarly, external monitoring specialists are sometimes former staff. Background knowledge also suggests that the kind of impact assessed depends on the individual evaluators, their skills, preferences, or biases.

Who conducts the study has a significant effect on the results. (1) *Internal project staff* are expected to have the greatest knowledge and access to data sources but also the greatest bias. They operate with very limited resources and time to conduct such studies and primarily follow guidelines provided by donors or internal documents. For that reason, these guiding documents tend to be short and actionable, and such evaluations can hardly apply any but the simplest methodologies. Generally, PMs or project staff are biased to look for outcomes and impacts that are in line with the project *logical frame*. Due to the specificity of the project's scope, even they doubt the usefulness of such evaluations, as impact, by definition, goes beyond the scope of specific projects or interventions.

The (2) *specialist internal monitoring staff* is the most diverse category, as in some cases, these are involved in the evaluations quite intimately, conducting in-person data gathering or leading focus groups and seminars, while in other cases, they act more as advisors and engage mostly in desk review of project materials. The discussion about the usefulness and applicability of evaluations is of greatest importance to this group and will be discussed in the following chapter in detail, but these evaluators, unless fulfilling a specific donor requirement, focus mostly on the usefulness of information gathered in the evaluation and institutional learning. In other words, they are biased to produce findings that can be transferred to other projects across the organization or to evaluate the effectiveness of the organization's operation.

With (3) *external specialists,* the quality of the results can vary significantly, and results are dependent on the resources and scope of the commission. On one hand, external evaluators can be former NGO staff or advisers, somewhat familiar with the project specifics; on the other, there are renowned expert companies (like the aforementioned 3ie) or universities and academics, generally more well-versed in more complex, sophisticated methodologies required for impact assessment. As a result, the latter group is far more careful with the definition of impact and what can truly be observed as an effect of the project interventions. But this comes at a price of greater separation between them and the practitioners. Interviewees often mentioned frustration with the limited applicability of such results and the difficulty in communicating the often complicated and nuanced results to staff who are not familiar with such methodologies. In the case of the former group, interviewees were at times voicing frustration and sometimes skepticism, as they would occasionally receive reports of dubious quality. In general, external evaluators are not so closely familiar with the

projects or the organization's functioning, so they can more easily miss the mark and produce findings that are too general and/or not applicable.

3.2. Key Topic Clusters

3.2.1. Measuring impact vs. improving quality

Among the key topics emerging from the analysis turned out to be the utility of impact evaluation. In practice, measuring and evaluating are usually integrated within a broader department that is also responsible for monitoring, learning, and quality control[11]. Moreover, given the *number of resources* needed—in the form of time and attention, both of staff and beneficiaries, money, and effort[12]—and, in line with the approximation of NGOs and regular businesses, organizations are naturally more focused on their efficiency and constant 'improvement.' The increasing focus on this type of monitoring and evaluation, and more generally towards results-based management, is a clear and ongoing trend. Language like "adaptive management," "results-based management," and other NPM-influenced language is being routinely used by the interviewees.

Therefore, the main reservation against impact studies that more rigorously focus on the long-term impacts of specific interventions is the perceived lack of utility of such practice. Measuring impact this way is seen as too distant and impractical. To justify such contributions, the institutions expect maximum utility and applicability of its outcomes channeled into future projects and strategies going forward (see footnote 6). In other words, from an organizational perspective, only findings that are 'useful' in serving future projects are worth doing, with a preference for actionable, short-term findings over long-term, difficult-to-untangle, and apply impacts[13].

This trend to prefer practical and transferable learning-oriented studies was present across cases, with a clear rationale that 'better-led projects, improving based on continuous delivery of information, and result-based management will have a better chance to deliver desirable, long-term impacts.' In such studies, regularity and continuity beat methodology, which is seen as the least important[14]. They are guided by desired lessons, which makes the format of

---

[11] In one case, the monitoring and evaluation department was separate from technical, that would handle the overseeing of expernal researchers conducting impact assessments. But still, within the organization, their importance was minimal compared to the monitoring and learning function.

[12] Interviewees from Case 2 would use expressions like: *"it's not about research, it's about performance"*, or *"we focused on evaluations as a learning exercise,"* while interviewees in Case 3 mentioned that: *"it's very difficult, obviously, to fund these. They're very expensive and demand so many resources, not just in terms of funding, but also on our side, and from the community that we would be working in. So, they're quite challenging to implement."* Interviewee in Case 1 mentioned: *"we need to be clear when we do something like this, because it's not just funding. It also means taking people's attention."*

[13] Interviewees were mentioning that the team *"would not understand the statistical logic or econometric analysis" (...) "let alone share it with others or discuss"* or that: *"[We] don't want a report that's 100 pages that nobody will have time to read. We want it to be quite succinct and actually get to the point."*

[14] Regarding learning-oriented evaluations, interviewee in Case 2 mentioned: *"for us, it does not matter so much the method as long as you have the time to bring people together to reflect on progress and on the results of your project. So for us, the method is the least [important]"*.

very clearly defined questions that are answered by the study in the most actionable way very applicable. Even when ex-post impact evaluations or similar studies were conducted, sound methodology (i.e. random sampling) gave way to practicality (i.e., zooming in on areas of future importance).

This practice could also be seen as continual testing and adjusting the logic of the intervention's logical framework. The underlying knowledge here is that this short-term learning and improving incremental links within the logical framework will lead to change and impact in the end, even without necessarily measuring or waiting to see whether this will be the case[15]. The shortcomings of this approach, and of not systematically and soundly measuring impact in general, are acknowledged[16]. This practice is in line with warnings in the literature about the overreliance on the notion of linearity of impact, as previously discussed (see 1.5 and Belcher, Palenberg 2018, Chianca 2008).

With regard to usefulness, the quality of the externally conducted impact assessments, which would be the preferred kind due to methodological and practical requirements and the need for specialist expertise, was also mentioned as a key obstacle. Bridging the organizational needs, understanding project particularities, and the ease of translating into practice would be used as an argument against such studies and for moving more towards more practical, quality improvement-oriented studies[17].

3.2.2. Time

Impact measuring takes time. Data gathering has to occur before and during the project but, most importantly, *ex post* with sufficient time for any long-term effects to manifest. For the studied NGOs, which predominantly operate in the short-term project cycles, this is very impractical, as with such temporal separation, either the projects already concluded in the area, the focus of local projects has changed, donor priorities changed or the way of approaching development within their project has evolved[18]. In other words, whether the impact evaluation study finds out if such a project was effective in delivering its desired impact or not, by now, there is no project to improve based on these findings. The

---

[15] Interviewee in Case 1 mentioning: *"... [I]t's kind of difficult to go very much into the direction of impact when you're consolidating data across different projects. So this is more from the point of view of project efficiency, with the view that that's contributing then to impact. The logic being that if we do our work better, with higher quality, then that also contributes to impact."*

[16] Interviewee in Case 1 mentioning: *"[W]hen it comes to the connection between outcomes and impact, then you see the assumptions are quite vague. For example, the assumption being that the political situation in the country will remain stable."*

[17] As interviewee in Case 3 put it: *"It's not necessarily about the fact that recommendations would be negative or anything like that because we're very interested in receiving those. It's more about the quality of the data, the fact that the recommendations come back very vague and not particularly useful, or just a lack of understanding of challenges in the sector and not taking that into account".*

[18] Interviewee in Case 3 mentioning: *"We've been trying to add as many ex-post evaluations as possible, which would basically come back to an area, let's say, a year later, or even a few years later (...) Obviously, donors are not necessarily willing to allow that, or they won't fund a few years later to come back to do an evaluation."*

organization might know if it was effective in the past but cannot do anything about it. These facts are overtly stated by all interviewees and are routinely mentioned in their internal documents.

The separation between the gradual, long-term nature of impact to fully manifest and the ever-increasing pace of the operation of these organizations could be seen as another factor pushing them towards a focus on continuous monitoring practices and building feedback loops that are, on an ongoing basis, feeding into the project planning and management, over impact evaluations that, by definition, require a step back and sufficient temporal separation. This could be further supported by the fact that when applying for short-term projects, donors would require data in order to continue cooperation well before any long-term impacts could realistically manifest.

In specific instances of long-term cooperation with donors on the same or similar projects, and/or developing a longer-term strategy of their approach, more space opens up for impact evaluations[19]. These were the stated instances of when the rare impact evaluations would be conducted. There seems to be a trend among some donors to switch to longer project terms in order to allow for more comprehensive effects and, in turn, also for more in-depth impact evaluations[20]. Time also links to the perceived utility and value-added of impact evaluations, as time for data collection, analysis, delivery, and integration by key recipients of the reports seems to be a growing priority.

3.2.3. Funding

As expected, funding is a main consideration when performing impact evaluations. The dominant knowledge is that impact evaluations are too expensive for their potential utility (or lack thereof)[21]. The frequently voiced issue was the difficulty of receiving funding within the short-term project cycle, as it is difficult to perform ex-post studies when the project implementation period is over and there is no funding, even though this is the defining feature of impact measuring.

Three strategies are being employed to mitigate this issue:

1. Resource pooling with different projects that cover the same topic or region and would overlap in any other way. This allows the funds dedicated to

---

[19] Interviewee in Case 2 outlining: "*With our main donor, which is [development agency], we think in phases. We usually have three to four phases. That means we are talking about nine to twelve years (...) and usually the donor will ask for an external evaluation at the end of the phase.*" Adding that in order to fund impact assessments, money needs to be fundraised from somewhere, and "*we can do it if we can convince (development agency) or another donor to say, 'okay, let's do an impact assessment.'*"

[20] An interviewee from Case 3 mentioned this specific programme taking place, allowing for community-led decision making and impact measurement, stating that: "*it was for six years and now the new (project) will actually be, in the new phase, for 12 years (...) it's the approach of the donor specifically (...) they have developed a tool for measuring resilience (...) And then you come back every few years to apply the same tool and basically see how the overall level of resilience is changing.*"

[21] These sentiments were repeated throughout internal documents and mentioned by the majority of interviewees.

monitoring and evaluation in one ongoing project to be used also to track more long-term impacts related to previous interventions[22].

2. Some donors allow for or even require *ex post* impact evaluations of interventions they financed for predetermined funds provided or a number of projects implemented. Although this is rare.

3. The NGOs' own funds (usually individually fundraised via donations from the public or coming from their own revenues) can be used. However, these funds are typically very precious[23], ranging usually from 10-20% of the NGOs' total annual income, and are usually the only funds that are not tied to specific project activities and thus can be used for this type of evaluation (Concern Worldwide 2024, Helvetas 2024, People in Need 2024).

The high price of impact evaluations compared to their limited utility would be the main retreating concern and reason for their low number[24]. This was further underlined by the varying quality of externally conducted impact evaluations, described as almost a luck of the draw[25]. The vagueness and probabilistic nature of such evaluations were also cited as why they are not worth the funds spent on them. The obvious underlying knowledge here is that if money was not a question, the organizations would be very interested in closely measuring the impact of their work[26], but when they are forced to prioritize, other strategies, like sound project design, monitoring, or having more capable staff will deliver greater.

The impact of donor pressures, accountability to donors over other stakeholders, and the reorientations to donor requirements as part of the way development NGOs run today is well documented in the literature, and the analyzed materials were in line with the assumptions based on the literature review. Donor requirements as the main basis for how, when, or by whom an evaluation is conducted are routinely mentioned throughout internal documents and were mentioned countless times by all interviewees. Understandably, the dominant approach seems to be that any internal reporting guidelines or approaches are overruled by

---

[22] In the words of Interviewee in Case 1. *"where we have maybe three projects that are doing similar activities maybe in different areas or from different donors, then we can kind of pull together the resources and do one evaluation for all of them."*

[23] Interviewee in Case 2 outlines the situation as: *"If you want to do an impact assessment, that needs to come with a source, someone who'll pay for it (...) we wanted to do it every year but this (funding) must come from our head office fonds. That was not very attractive to us."*

[24] Interviewee in Case 3 summarising this concern as: *"if we find that after spending 20 or 30 thousand [per study], and we still can't learn from the experience, it's becoming quite difficult to appreciate those opportunities."*

[25] In the words of Interviewee in Case 3: *"we've been really struggling with the quality of the evaluations themselves, particularly when we hire external consultants."* This concern was then directly tied to to topic of funding: *"...you either need to allocate quite a lot of funding to ensure you get a consultant of good quality, or one who is very experienced and will then produce a quality report."* ... *"and even if you budget quite a big amount of funding toward that, you might still be unlucky, and the report might not be of good quality."*

[26] In the words of interviewee in Case 2: *"...we wanted to do this rigorous impact assessment, these randomized control trials and so forth, but due to the lack of funds and expertise, we gave up on that."*

specific donor requirements[27]. Given that these organizations are overwhelmingly dependent on project-based donor funding for their operation, this is the only logical way.

3.2.4. The DAC Criteria: outputs, outcomes, and impact

The majority of evaluation reports produced in all organizations are end-of-project evaluations[28]. These are often donor-requested and, if not, mandated by internal guidelines. In all cases, the OSCE DAC criteria are employed within these and are mandated by either internal guidelines or donor requirements. As outlined in Chapter 3, the DAC criteria also include "impact" with a very ambitious definition, especially for the context in which these are employed. As a result, it is commonplace that what are, by definition, outputs or outcomes of that particular project labeled as impact.

This confusion could be coming from a lack of methodological rigor, time constraints, or lack of training, but it most likely points to the limited usability of the DAC criteria, and especially the impact criterion, to serve this purpose. An end-of-project evaluation, almost by definition, cannot systematically assess the long-term changes—positive or negative, intended or unintended, and primary or secondary—that result from a development intervention, as defined by OECD (2023). A typical end-of-project evaluation comes nowhere near the methodological rigor necessary to track impact, as it is mostly produced by the project manager (PM) with little or no additional resources. As outlined in Chapter 1.5., by definition, "impact" would need to be measured with an *ex post* analysis, ideally employing some kind of counterfactual, an element of randomization, external evaluators, and observing some macro indicators. Due to the lack of resources for measurement, impact chapters in such reports often rely on anecdotal evidence or observations of short-term outcomes.

Since end-of-project evaluations employ the DAC criteria, the project outputs or outcomes are often, willingly or unwillingly, mislabelled as impact. Such evaluations frequently list project activities and outcomes in impact chapters rather than actual measurable impacts. This is sometimes accompanied by speculative impact estimations, using language like "could lead to," "should result in," "is expected to," or "may have contributed," when tying the project outputs or outcomes to impact, usually following the proposed logic from the project's logical frameworks or tables based on the theory of change, but lacking any sound measurements or evidence. Worryingly for impact assessment, this practice then solidifies the initial assumptions about the impact of development interventions (made in the planning

---

[27] As an example, one of the Cases boasts a decentralized working culture manifested by limited power of the headquarters over the conduct of individual country branches. In that case, the only mandatory practices are *de facto* mandated by the donors.

[28] It is a standard practice for longer projects or those with larger budgets to also require mid-term evaluations but project without final evaluations are very rare.

phase) without critically questioning, tracking, observing, or measuring whether this is, in fact, the case. The logic of impact is often drawn out but rarely actually measured.

The implicit knowledge about the use of the OECD DAC is mixed. It is considered the most classic way of approaching project evaluations, and the sentiments suggest that it is so embedded that any changes might be difficult. However, even though these are so heavily promoted, all of the interviewees had reservations about their use[29]. The criteria are often selected without much thought about what actually is the aim of the study or which questions should be answered. There is a common sentiment that the DAC criteria should be used as a tool and adapted more to the specific context, that they should be more user-friendly, and the overall approach more malleable. Still, however, the fact that they are used in contexts that do not permit the study of impacts is not being questioned.

The users realize that the individual criteria are often confused or interchanged. Some organizations have a way to expand the DAC by other approaches to fit them better into the desired contexts. Overall, the use of the DAC criteria as they are, with an implicit expectation to apply all of the criteria every time, without considering their usefulness or particular constraints (whether financial, temporal, or definitional), exists in a limbo between explicit and implicit expectations from the donor and other engaged organizations and between adaptability and rigidity. The inclusion of impact within these criteria is hailed by some and questioned by others.

The impact criterion is not the only problematic one from DAC, but its measurement is arguably the most problematic to end-users. As such, the term "impact" suffers from a double definition[30]—casually meaning the whole sum of effects and activities of a project or intervention, and as a more precise definition as stated in the OECD[31]. This confusion then leads, from one side, to the natural expectation to include impact in all evaluations, since it is the thing that everybody strives for, and perhaps even contributing to a myth that impact is actually measured in studies that are not suitable[32]. On the other hand, there is a

---

[29] For example, interviewee in Case 1 mentioning: "In my experience, most evaluations mention all the criteria, but the focus should be on what we aim to gain from the evaluation, not just following the criteria blindly."

[30] For example, interviewee in Case 1 outlined the confusion as: "...impact can mean many things and not only reflect on how many people received [*the project deliverable*] in that case (...) it's more on people leading the evaluation to identify what kind of impact, like, is it as per the project indicators? Or is it more like value for money, which is more efficient? But even impact in terms of what was not covered by regular monitoring, for instance, or the impact of a specific approach that we use in this project compared to any other approach."

[31] "Long-term changes—positive or negative, intended or unintended, and primary or secondary—that result from a development intervention (OECD 2023).

[32] Interviewee in Case 1 claiming, for example: "I would say impact is always there. Whatever criteria we prioritize or don't prioritize, impact is always there, especially in final evaluations."

notable frustration and powerlessness of PMs and other staff tasked with measuring what they are not able to measure[33].

3.2.5. Indicators

The selection of indicators used to measure impact illustrates the reality of practical impact measurement during most routine evaluations. To aid project evaluations and allow for broader comparisons[34], staff are sometimes given access to sample indicator banks to be used to measure individual DAC criteria for the particular project area. An overview of these shows that true long-term impact indicators are few and far between. In practice, when reviewing sample evaluations, sometimes a different indicator, not meant for measuring impact but outcomes, was used in the impact section even when an impact indicator was available in the indicator bank[35].

It must be noted that a common strategy in assessing impact in studies that did not have the necessary scope and resources to attempt to measure actual impact was the use of proxy indicators. This practice seems to be a common compensation strategy known among the staff responsible[36]. It further adds to the broader picture of using probability language and best guesses in impact assessments in place of more methodical measurements or sound evaluation.

The ambitiousness of using proper long-term indicators for impact in typical evaluations might be further illustrated by the fact that when a sufficiently funded broad impact evaluation fulfilled by external contractors was reviewed, it was much more aware of the difficulty in measuring certain impact indicators and in measuring impact in general. This resulted in this specialized and broad study using less ambitious and comprehensive impact indicators than much briefer and less comprehensive end-of-project evaluations would use and being much more careful in assessing what the true impact of project interventions was. Such evaluations would use less definite or unequivocal links between project interventions and long-term impacts and, even though they would generally still lean towards emphasizing linear causality, would be much more careful to attribute changes in simple cause-effect relationships.

---

[33] Another interviewee from Case 1 mentioning: "*To me, impact should not be related only to the project itself but also in a broader way to identify what are the side impacts, for instance, that we haven't been analyzing before because we don't have the capacity to evaluate everything all the time.*"

[34] That being said, comparisons of indicators across projects don't seem to be working as expected Interviewee in Case 1 noted that: "*We used to have something called global indicator, under each sector. They were basically indicators that we wanted to measure in as many of the projects as possible from that sector. It was not really working so well.*"

[35] No explanation was found for this instance, but given the time and funding constraints illustrated earlier, and lack of methodological expertise, this might have been motivated by practical feasibility concerns.

[36] As interviewee in Case 1 mentioned, "*For long-term impacts, we sometimes use proxy indicators or look at intermediate outcomes that can reasonably be expected to lead to longer-term changes.*" Adding that, "*[w]hile we couldn't measure the actual [*desired impact*] within the project timeframe, the [*proxy indicators*] gave us a good indication of potential future impacts.*"

3.2.6. Audience

Another key consideration is the intended audience of any evaluation, as it helps uncover the accountability rationale of the NGO staff conducting such evaluations at any time. With somewhat looser definitions and methodologies, the process of evaluation, inevitably, assumes elements of a performative exercise and the rationale of utility, which, however, does not equal bad intention. Hints or language artifacts of this rationale appeared throughout the materials and interviews[37]. While impact intertwines throughout the different kinds of reporting, it is only very sparsely the main goal of an evaluation.

In reality, this could manifest in ways such as that:

- if the intended audience is internal staff and the organization, the main goal, and therefore the inherent bias, could be towards of actionability of the evaluation findings and their ability to be integrated into the processes and planning. In this case, findings that are complex, particular, nuanced, or inconclusive might be discouraged.

- if the intended audience is local stakeholders, the main goal, and therefore the inherent bias could be towards accessibility[38], facilitating clear communication and good relations with local stakeholders that allow for future functioning.

- if the intended audience is donors and foundations, the main goal and therefore the inherent bias could be towards demonstrating positive outcomes, return on investment, efficiency, and sound project design that allows for future funding from that organization.

- if the intended audience is the general public, the main goal and therefore the inherent bias could be towards simplicity, clear storytelling, or showcasing heartwarming stories and visuals that hint at impact without the added complexity or ambiguity of impact measurement[39].

In any case, although all of these audiences could reasonably care for actual impact, it is never the main goal of reporting for them.

---

[37] For example, interviewee in Case 1 mentioned, "*We create different versions of the report to suit the needs of each audience.*"

[38] Interviewee in Case 1 mentioned that: "*For donors and management, we provide detailed reports with comprehensive data and analysis. For local communities and other stakeholders, we prepare summary reports or presentations that highlight key findings and recommendations in an accessible format.*"

[39] As one reviewed internal document puts it, "*Case studies can be a valuable addition to M&E and reporting, by illustrating complexities that may be hard to capture with other means. They can help to put a human face to the numbers that reflect progress towards achieving impact indicators, making reports and visibility materials more interesting.*"

3.3. Discussion of Results

Seeing how international development NGOs approach and report impact a invites a broader discussion about the role of such evaluations and reporting in their operations and possible ways forward. There is no doubt that measuring true impact is very difficult. Studies that could even attempt gauging what the impact of a particular intervention was would need to zoom out and step back very much, so much in fact, that they would become quite disconnected from the project realities and everyday operations of the involved organizations. This, in combination with the methodological complexity, is evident from the practice of impact assessment being done by external contractors. This invites a practical consideration of whether it should be primarily the NGOs themselves in the first place that conduct the impact assessments, or whether it should not be the responsibility of someone else, perhaps the donor organizations, that would feed back into the operations of the NGOs.

There is no doubt that at their core, all of the studied organizations wish for impact, but many of the operational, financial, temporal, and other factors prevent them from measuring it to the full extent possible. On top of this already challenging situation, a host of methodological hurdles are stacked. As such, the doings of these organizations or projects are very difficult to isolate from a host of outside variables that are influencing any given situation at any given time. The closer one comes to uncovering the impact in the particular situation, the more unique and particular such findings would be. Therefore, almost by definition, impact assessment findings operate on a spectrum between accuracy and usability, with usability (followed by actionability) winning out for the studied organizations at this point. The observed NGOs are not academic institutions but practitioners; this outcome is understandable and only adds to the question of whether it should be the NGOs who study the impacts of their work.

Also, since their actions are only a part of the broader development discourse, its impacts should be interpreted as such. Many of the mentioned constraints can be influenced by the NGOs only to a limited extent—for instance, they could vouch for extended project periods that could integrate extended time for *ex post* evaluation—but no single organization, especially mid- or small-sized, could transform the short-term project cycle as a norm in the sector. The findings of impact assessments should, thus, integrate such limitations into their considerations, and as such, their findings would then transcend the operations of a single NGO. Taking this reality into consideration and understanding the inherent lack of resources—both monetary and other—could result in impact assessment being taken out of the realm of these organizations. In reality, this could be done, for example, by pooling resources between organizations active in a particular area (topically, temporally, and/or geographically) and commissioning joint, broader impact evaluations.

Moreover, *impact* and *measuring impact* are concepts that are definitionally extensions of the Western-centric positivist thinking of the past few centuries[40]. It is understandable that organizations rooted in this cultural tradition perceive it as an important aspect of their work. However, as the trend of continual inclusion of local stakeholders in the operations of these organizations continues, positivist practices such as these might start losing their relevance or become influenced by competing worldviews and approaches to functioning within the world. As such, the concept of impact might start evolving too, or perhaps the notion of the impact of any man-made organizations in the world might start *correcting* back to offer a more balanced and representative outlook.

Following up on the concerns found in the literature on the professionalization and bureaucratization of NGOs (Chapter 1.3.), the above-presented findings appear to support the fears associated with the adoption of formal management techniques, especially that the adoption of the approaches of the New Public Management (NPM) could lead to the reduction of complex social issues into technical problems (as illustrated by the preference for monitoring that solidifies the logical frame assumptions), thereby limiting their ability to address the root causes of poverty and inequality (Golini et al 2015; Hwang & Powell 2009).

The topic of donor pressures was ever-present throughout the analysis, hinting that the need to serve donor interests was a big part of the implicit knowledge guiding the behavior of NGOs in impact reporting. The discrepancy between the lofty ideals and ambitious impacts outlined in the organizations' strategic documents and the practical guidelines, which disproportionally remind and warn about donor requirements above anything else, could perhaps be the most prominent proxy indicator for such pressure. That being said, the nature of such pressure, sources for donor accountability over accountability to other stakeholders, the overt or covert exercise of state or corporate interests through NGO programs, or other elements of this relationship, could not be directly investigated within this design. This topic, beyond structural limitations, is very sensitive and not easily observable. Reconstructing background knowledge related to donor policy and pressure perceptions, or potentially uncovering the background knowledge in cases of clash between stakeholder and donor interests within the NGO space, could present an interesting natural topic for further inquiry.

However, donor pressure can be viewed from another side, too. The growing focus on measuring short- or medium-term, measurable outcomes over long-term impact was clear. However, the source of such pressures might not be solely attributed to donor requirements, as perhaps stated in the literature described in Chapter 1.3. The impact of the NGOs' own operational interests and desire to improve the efficiency and project planning process appeared more often and quite overtly in the observed cases. The lack of operational utility of comprehensive impact assessment seemed to be as relevant as other aspects, like the lack of funding of the short-term project cycle limitations on the negligible amounts of *ex post* impact assessments.

---

[40] As illustrated previously on the research of Meyer (2014).

Conclusion

This analysis presented a unique look into the inner functioning and practices of development NGOs that grew to become important and influential organizations over the past decades. Boasting budgets in the hundreds of millions of US dollars, employing thousands of highly skilled individuals, and operating across the globe, international development NGOs are a fascinating feature of international relations realm. The degree of professionalization also results in considerable amounts of administrative work produced, a significant part of which is spent on reporting and monitoring.

The pervasive use of evaluations of different kinds is certainly impressive, hinting at a considerable portion of time of the employees being spent on reporting, yet the limited actual impact measurement is similarly staggering. The analysis found that only around one percent of implemented projects undergo a methodologically sound *ex post* impact assessment utilizing elements of random control trials or counterfactuals. And these are still only selected flagship projects, further decreasing the statistical representativeness of such monitoring.

Donor requirements, lack of disposable funds for impact assessment, and the reality of the short-term project cycle are the chief unifying factors of the examined cases. But despite these similarities, the examined organizations' internal structures and cultures lead to some diversity in how evaluations are designed and applied. Donors remain the strongest influence: if a donor mandates a certain evaluation form, the organization typically follows it. However, in the absence of donor requirements, organizations may opt for alternative methods. Despite that, the factors discouraging impact assessments—their cost, demanding methodologies, limited actionability and usability of their results, and expertise required— still prevail as the organizations use different forms of monitoring and evaluation instead.

The implicit knowledge guiding impact evaluation is anchored in several overlapping assumptions and organizational norms. Short project cycles and donor-driven requirements lead staff to prioritize easily measured outcomes over long-term outcomes, reinforcing the belief that continuous, results-based management will ultimately deliver impact without formally verifying it. These time-bound funding structures also foster the sense that rigorous *ex-post* studies are financially and operationally prohibitive, thus implicitly accepting that in-depth impact assessment is best reserved for exceptional flagship projects or special donor mandates. The NGOs' internal culture is also shaped by the conviction that methodical precision often yields limited practical returns. Staff perceive in-depth studies—particularly when externally commissioned and highly technical—as difficult to integrate into everyday program adaptations. Consequently, evaluation practices emphasize actionability and immediate utility, relying on outcome indicators as proxies for impact. This collectively sustains a tacit consensus that, while genuine impact matters, concretely measuring it is secondary to maintaining donor compliance, refining implementation, and preserving operational momentum.

The ongoing move from more traditional, impact-focused studies towards results-based

management and continuous monitoring and project refinement instead was certainly an observable trend. In hindsight, this approach is understandable, especially due to improved utility, resource constraints, and short project cycles. However, only careful, continuous assessment will tell whether such an approach can, indeed, replace long-term impact evaluation.

The continuous prevalence of the OECD DAC Criteria might represent the biggest hurdle to genuine impact assessment. The fact that these criteria are so ubiquitous in so many forms of evaluations across the work of all of the NGOs, combined with the fact that they include "impact" as one of those criteria, contributes to forming a dangerous illusion. Certainly, the NGO staff realizes the shortcomings of these indicators, especially the fact that in most cases when it is used, the impact cannot be actually measured, but an external observer could be mistakenly made to believe that impact is in fact analyzed extensively, on every step of the way. This myth is further aided by the confusion between indicators in this set—especially outputs, outcomes, and impact—and the loose, probabilistic nature in which it is commonly assessed in reports.

Impact is a tricky thing. Although everyone would like to have a positive impact, very few are truly motivated and given the opportunity to measure it, as its actual measuring is very difficult, messy, rarely straightforward, difficult to generalize, and uncomfortable.

Several interesting questions come out of this study, chiefly:

1. Should NGOs themselves be responsible for fully fledged impact evaluations, or should donors or third parties (academia, multi-NGO consortiums) undertake them?

2. Given that long-term impacts are influenced by a host of other factors out of the NGOs' control (be it structural constraints of the short-term project cycle or outside sociopolitical and other factors), what informational and actionable value would their measuring truly offer to the organizations to justify their high cost?

3. What is the actual aim of the number of bureaucratic and reporting hurdles that are weighted down onto the NGOs' shoulders when impact is so rarely measured?

In sum, while international development NGOs aspire to achieve lasting impact, practical realities—from funding and timing constraints to donor-driven short-term evaluations—encourage them to focus on incremental learning and immediate operational improvements. True impact is often assumed rather than conclusively measured, leaving open the question of how (and by whom) rigorous, long-term impact evaluations should be conducted.

Závěr

Prezentovaná analýza nabízí unikátní vhled do praktického interního fungování mezinárodních nevládních organizací v rozvojovém sektoru, ze kterých se v posledních

dekádách staly vlivné organizace. Operují s rozpočty v řádech stovek milionů dolarů ročně, zaměstnávají tisíce vysoce kvalifikovaných zaměstnanců a operují po celém světě. Tyto organizace jsou bezpochyby impozantními příklady lidské vynalézavosti. Míra jejich profesionalizace také vede k produkci velikého množství administrativních výstupů, z nichž podstatná část zahrnuje monitoring a vyhodnocování.

Takto rozsáhlá produkce velkého množství různých druhů evaluací, i když bezpochyby působivá, poukazuje na značné množství času, který této činnosti zaměstnanci věnují; neméně pozoruhodné je však i to, že jak málo se ve skutečnosti měří impakt. Tato analýza poukázala na to, že jen asi jedno procento realizovaných projektů prochází metodologicky rigorózními *ex post* evaluacemi impaktu, která by využívaly přístupy randomizovaných kontrolních studií nebo kontrafaktuální metody. A i v těchto omezených případech se jedná zpravidla o vybrané, tzv. vlajkové projekty, což dále snižuje reprezentativnost takto pojatého sledování.

Klíčové podobnosti sledovaných případů ve vztahu k hodnocení impaktů byly požadavky donorů, nedostatek vlastních disponibilních prostředků k financování těchto evaluací a realita fungování v rámci krátkodobých projektových cyklů. I navzdory těmto podobnostem se ale interní fungování a profesní kultury zkoumaných organizací liší, což vede i k rozdílným přístupům, se kterým evaluace realizují. I tak ale mají donoři na tuto praxi rozhodující vliv. Pokud donorská organizace vyžaduje určitý způsob evaluací, organizace jim obvykle vyhoví. Pokud ale donoři podobné požadavky nemají, sahají organizace po odlišných metodách. I v takovém případě ale převažují faktory komplikující komplexní hodnocení impaktu. Z důvodu, že podobné studie s sebou nesou vysoké finanční náklady, vyžadují náročné metodologické postupy a expertízu, , a jejich výstupy jsou organizacemi omezeně využitelné, nakonec organizace volí jiné způsoby evaluace.

Hodnocení impaktu v těchto organizacích ovlivňuje řada implicitních znalostí vycházeních z několika vzájemně provázaných předpokladů a organizačních zvyklostí. krátkodobé projektové cykly a požadavky donorů vedou k upřednostňování snadno měřitelných výstupů před dlouhodobými impakty, což posiluje přesvědčení, že průběžné sledování a výsledky informovaný management (RBM) samy o sobě zaručí dosažení impaktu i bez formálního ověřování a sledování. Krátkodobé formy financování navíc ještě upevňují dojem, že důkladné *ex post* studie impaktu jsou z hlediska financí i provozu příliš náročné, a měly by se tak týkat spíše výjimečných, tzv. vlajkových projektů nebo vyhovění zvláštních požadavků ze strany donorů. Interní organizační kultura je ovlivněna i názorem, že vysoce metodologicky propracované studie často nenabízejí dostatečně praktický přínos pro každodenní fungování. Zaměstnanci vnímají detailní, zejména externě zadávané a metodologicky náročné evaluace impaktu jako obtížně použitelné pro každodenní řízení projektů. Z toho důvodu se v evaluacích klade důraz zejména na praktičnost a okamžitou využitelnost, přičemž indikátory výstupů z projektů (*outcomes*) často slouží jako náhražka přímého měření skutečného impaktu (*impacts*).

V rámci zkoumaných případů byl jednoznačně patrný odklon od důkladných studií

zaměřených zkoumajících impakt směrem k průběžnému sledování výsledků a tzv. adaptivnímu řízení projektů. S ohledem na omezené zdroje, krátké trvání projektů a požadavek na rychlou zpětnou vazbu je tato tendence pochopitelná. Otázkou však zůstává, zda takový přístup může skutečně nahradit evaluace zaměřené na sledování dlouhodobého impakt. Pouze hlubší a opakované sledování časem může ukázat, jestli lze touto cestou spolehlivě podchytit a předpokládat, jaký dlouhodobý impakt projekty skutečně mají.

Možná největší překážkou pro důsledné měření impaktu je zásadní role kritérií OSCE DAC.

Tato kritéria jsou využívána v celé škále evaluací napříč všemi zkoumanými organizacemi, a jelikož přímo zahrnují i kritérium „impaktu," vzniká klamný dojem, že je impakt na každém kroku pečlivě hodnocen. Ačkoli samotní zaměstnanci si uvědomují slabiny spojené s těmito indikátory, zejména to, že v rámci většiny běžných evaluací nelze impakt skutečně zachytit, navenek může tato situace působit, jako byl impakt plošně a detailně sledován napříč projekty. Tento dojem navíc ještě umocňuje zaměňování ukazatelů výstupů, výsledků a impaktu (*outputs, outcomes and impact*), společně s nejednoznačným, spíše pravděpodobnostním jazykem, který se v hodnoceních „impaktu" objevuje.

Samotný impakt představuje komplexní fenomén. Každý by rád dosáhl žádaného pozitivního impaktu, jen málokdo ale jeho dosahování skutečně měří a nebo k tomuto měření má dostatečné zdroje, jelikož jeho reálné měření je složité, nejednoznačné, jen zřídka přímočaré, špatně přenositelné do jiných kontextů a ne nutně komfortní.

Ze studie vyplývá několik dalších, zásadních otázek:

1. Měly by být tyto organizace samy vůbec odpovědné za provádění plnohodnotného mření impaktu, nebo by tuto roli měly převzít donorské organizace či nezávislé subjekty (například akademická sféra či konsorcia několika různých neziskových organizací)?
2. Vzhledem k tomu, že dlouhodobý impakt ovlivňuje mnoho vnějších faktorů, které organizace nemají plně pod kontrolou (např. omezení krátkodobých projektových cyklů nebo sociopolitické okolnosti), je na místě otázka, jakou reálnou informační hodnotu pro ně takové vyhodnocení má, zejména pokud má ospravedlnit vynaložení takto vysokých nákladů?
3. Jaký je skutečný účel bezpočtu administrativních a reportovacích požadavků kladených na organizace, když impakt se skutečně měří jen velice zřídka?

Celkově vzato, ačkoli se mezinárodní rozvojové nevládní organizace snaží dlouhodobě dosáhnout pozitivního impaktu, praktická omezení – od financování a krátkodobých projektových cyklů, po donory upřednostňované krátkodobé evaluace – je vedou spíše k průběžnému zlepšování, zaměřenému na operativní vylepšování probíhajících intervencí. Skutečný impakt se pak v tomto procesu většinou pouze předpokládá, místo aby byl skutečně sledován a měřen. Zbývá tak otázka, jakým způsobem (a kým) by mělo být důkladné, dlouhodobě zaměřené hodnocení impaktu uskutečňováno.

List of References

ADLER, Emanuel and POULIOT, Vincent (eds.), 2011a. *International practices*. Cambridge : New York: Cambridge University Press. Cambridge studies in international relations, 119. ISBN 978-0-521-28117-1.

ADLER, Emanuel and POULIOT, Vincent, 2011b. International practices. *International Theory*. 18 February 2011. Vol. 3, no. 1, p. 1–36. DOI 10.1017/S175297191000031X.

ADLER, Emanuel, 2012. Constructivism in International Relations: Sources, Contributions, and Debates. In: *Handbook of International Relations*. Online. SAGE Publications. ISBN 978-1-4739-7119-6. Available from: <http://www.sagereference.com/hdbk_intlrelations/Article_n10.html>.

ADLER-NISSEN, Rebecca and POULIOT, Vincent, 2014. Power in practice: Negotiating the international intervention in Libya. *European Journal of International Relations*. December 2014. Vol. 20, no. 4, p. 889–911. DOI 10.1177/1354066113512702.

AHMED, Shamima and POTTER, David M., 2013. *NGOs in international politics*. Repr. Boulder, Colo: Kumarian Press. ISBN 978-1-56549-230-1.

ATACK, Iain, 1999. Four Criteria of Development NGO Legitimacy. *World Development*. May 1999. Vol. 27, no. 5, p. 855–864. DOI 10.1016/S0305-750X(99)00033-9.

BANKS, Nicola and BROCKINGTON, Dan, 2019. *Mapping the UK's development NGOs: income, geography and contributions to international development*. Manchester: The University of Manchester Global Development Institute. Working paper series / Global Development Institute, 035 (2019). ISBN 978-1-909336-70-4.

BANKS, Nicola and HULME, David, 2012. The Role of NGOs and Civil Society in Development and Poverty Reduction. *SSRN Electronic Journal*. Online. 2012. DOI 10.2139/ssrn.2072157.

BANKS, Nicola, HULME, David and EDWARDS, Michael, 2015. NGOs, States, and Donors Revisited: Still Too Close for Comfort? *World Development*. February 2015. Vol. 66, p. 707–718. DOI 10.1016/j.worlddev.2014.09.028.

BATLIWALA, Srilatha, 2007. Taking the power out of empowerment – an experiential account. *Development in Practice*. August 2007. Vol. 17, no. 4–5, p. 557–565. DOI 10.1080/09614520701469559.

BELCHER, Brian and PALENBERG, Markus, 2018. Outcomes and Impacts of Development Interventions: Toward Conceptual Clarity. *American Journal of Evaluation*. December 2018. Vol. 39, no. 4, p. 478–495. DOI 10.1177/1098214018765698.

BETH A. SIMMONS AND LISA L. MARTIN, [no date]. International Organizations and Institutions. In: *Handbook of International Relations*. Online. SAGE Publications. ISBN 978-1-4739-7119-6. Available from: <http://www.sagereference.com/hdbk_intlrelations/Article_n10.html>.

BOURDIEU, Pierre, 1977. *Outline of a theory of practice*. Cambridge: Cambridge University Press. Cambridge studies in social and cultural anthropology, 16. ISBN 978-0-511-81250-7.

BRASS, Jennifer N., LONGHOFER, Wesley, ROBINSON, Rachel S. and SCHNABLE, Allison, 2018. NGOs and international development: A review of thirty-five years of scholarship. *World Development*. December 2018. Vol. 112, p. 136–149. DOI 10.1016/j.worlddev.2018.07.016.

BRIÈRE, Sophie, PROULX, Denis, FLORES, Olga Navaro and LAPORTE, Mélissa, 2015. Competencies of project managers in international NGOs: Perceptions of practitioners. *International Journal of Project Management*. January 2015. Vol. 33, no. 1, p. 116–125. DOI 10.1016/j.ijproman.2014.04.010.

BUEGER, Christian, 2014. Pathways to practice: praxiography and international politics. *European Political Science Review*. August 2014. Vol. 6, no. 3, p. 383–406. DOI 10.1017/S1755773913000167.

CARLSNAES, Walter, RISSE, Thomas and SIMMONS, Beth, 2002. *Handbook of International Relations*. Online. 1 Oliver's Yard, 55 City Road, London EC1Y 1SP United Kingdom: SAGE Publications Ltd. ISBN 978-0-7619-6305-9. [Accessed 19 October 2024].

CHAE, Young-Gil, 2008. Development Discourse. In: DONSBACH, Wolfgang (ed.), *The International Encyclopedia of Communication*. Online. 1. Wiley. ISBN 978-1-4051-3199-5. [Accessed 12 November 2024].

CHAHIM, Dean and PRAKASH, Aseem, 2014. NGOization, Foreign Funding, and the Nicaraguan Civil Society. *Voluntas: International Journal of Voluntary and Nonprofit Organizations*. 2014. Vol. 25, no. 2, p. 487–513.

CHIANCA, Thomaz, 2008. The OECD/DAC Criteria for International Development Evaluations: An Assessment and Ideas for Improvement. *Journal of MultiDisciplinary Evaluation*. 29 February 2008. Vol. 5, no. 9, p. 41–51. DOI 10.56645/jmde.v5i9.167.

CONCERN WORLDWIDE, 2024. *Annual Report & Financial Statements 2023*. Online. Available from: https://admin.concern.net/sites/default/files/documents/2024-06/AR23_interactive.pdf .

COOK, Thomas D. and CAMPBELL, D.T., 1979. *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Houghton Mifflin.

CORNWALL, Andrea and BROCK, Karen, 2005. What Do Buzzwords Do for Development Policy? A Critical Look at "Participation", "Empowerment" and "Poverty Reduction." *Third World Quarterly*. 2005. Vol. 26, no. 7, p. 1043–1060.

COX, Robert W. (ed.), 1996. Social forces, states, and world orders: beyond international relations theory (1981). In: , *Approaches to World Order*. Online. Cambridge: Cambridge University Press. p. 85–123. Cambridge Studies in International Relations. ISBN 978-0-521-46651-6.

CRACK, Angela, 2019. NGO accountability. In: *Routledge handbook of NGOs and international relations*. London: Routledge, Taylor & Francis Group. ISBN 978-1-138-28550-7.

CRUSH, Jonathan (Ed.), 1995. *Power of Development*. London: Routledge. ISBN 978-0415113575.

DAVEY, Calum, AIKEN, Alexander M, HAYES, Richard J and HARGREAVES, James R, 2015. Re-analysis of health and educational impacts of a school-based deworming programme in western Kenya: a statistical replication of a cluster quasi-randomized stepped-wedge trial. *International Journal of Epidemiology*. October 2015. Vol. 44, no. 5, p. 1581–1592. DOI 10.1093/ije/dyv128.

DAVIES, Thomas Richard, 2014. *NGOs: a new history of transnational civil society*. Oxford: Oxford University Press. ISBN 978-0-19-023564-2.

DAVIS, John-Michael, 2019. Real "non-governmental" aid and poverty: comparing privately and

publicly financed NGOs in Canada. *Canadian Journal of Development Studies / Revue canadienne d'études du développement*. 3 July 2019. Vol. 40, no. 3, p. 369–386. DOI 10.1080/02255189.2019.1556623.

ECOSOC, 1996. *ECOSOC resolution 1996/31.*. Online. 1996. Available from: https://www.unov.org/documents/NGO/NGO_Resolution_1996_31.pdf

EDWARDS, Michael and HULME, David, 1996. Too close for comfort? the impact of official aid on nongovernmental organizations. *World Development*. June 1996. Vol. 24, no. 6, p. 961–973. DOI 10.1016/0305-750X(96)00019-8.

EDWARDS, Michael and HULME, David, 2014. *Non-Governmental Organisations - Performance and Accountability: Beyond the Magic Bullet*. Hoboken: Taylor and Francis. ISBN 978-1-85383-310-6.

ESCOBAR, Arturo, 1995. *Encountering development: the making and unmaking of the Third World*. Princeton, NJ: Princeton Univ. Press. Princeton studies in culture, power, history. ISBN 978-0-691-00102-9.

EYBEN, Rosalind, 2013. Uncovering the politics of 'evidence'and 'results': A framing paper for development practitioners. In: *Prepared for the Politics of Evidence Conference in Brighton, UK*. 2013.

FERGUSON, James, 1994. *The anti-politics machine: "development," depoliticization, and bureaucratic power in Lesotho*. Minneapolis: University of Minnesota Press. ISBN 978-0-8166-2437-9.

FOUCAULT, Michel, 2009. *Archaeology of knowledge*. 1st publ., repr. London, New York, NY: Routledge. Routledge classics. ISBN 978-0-415-28753-1.

FOUCAULT, Michel, 2019. *Discipline and punish: the birth of the prison*. London: Penguin Books. ISBN 978-0-14-199133-7.

FOWLER, Alan (ed.), 2011. Development NGOs. In: , *The Oxford Handbook of Civil Society*. Online. 1. Oxford University Press. ISBN 978-0-19-539857-1. [Accessed 18 October 2024].

GERTLER, Paul J., MARTINEZ, Sebastian, PREMAND, Patrick, RAWLINGS, Laura B. and VERMEERSCH, Christel M. J., 2016. *Impact Evaluation in Practice, Second Edition*. Online. Washington, DC: Inter-American Development Bank and World Bank. ISBN 978-1-4648-0779-4. [Accessed 26 November 2024].

GLEWWE, Paul and TODD, Petra, 2022. *Impact Evaluation in International Development: Theory, Methods, and Practice*. Washington, DC: World Bank Group. ISBN 978-1-4648-1498-3.

GOLINI, Ruggero, KALCHSCHMIDT, Matteo and LANDONI, Paolo, 2015. Adoption of project management practices: The impact on international development projects of non-governmental organizations. *International Journal of Project Management*. April 2015. Vol. 33, no. 3, p. 650–663. DOI 10.1016/j.ijproman.2014.09.006.

GOLSORKHI, Damon (ed.), 2010. *Cambridge handbook of strategy as practice*. Cambridge: Cambridge University Press. ISBN 978-0-511-77788-2.

GROSS STEIN, Janice, 2011. Background knowledge in the foreground. In: *International Practices*. Cambridge: Cambridge University Press. Cambridge Studies in International Relations.

ISBN 978-0-511-86237-3.

HAJER, Maarten A. and WAGENAAR, Hendrik (eds.), 2003. *Deliberative Policy Analysis: Understanding Governance in the Network Society*. Online. 1. Cambridge University Press. ISBN 978-0-521-53070-5. [Accessed 23 November 2024].

HARVEY, David, 2020. *A brief history of neoliberalism*. Oxford: Oxford University Press. Oxford scholarship online. ISBN 978-0-19-191766-0.

HEARN, Julie, 1998. The 'NGO-isation' of Kenyan society: USAID & the restructuring of health care. *Review of African Political Economy*. Online. 1 March 1998. Vol. 25, no. 75. DOI 10.1080/03056249808704294. [Accessed 17 October 2024].

HELLINGER, Stephen, HELLINGER, Douglas and O'REGAN, Fred, 1988. *Aid for just development: report on the future of foreign assistance*. 3. [print.]. Boulder: Rienner. ISBN 978-1-55587-122-2.

HELVETAS, 2024. *ANNUAL REPORT 2023*. Online. Available from: https://www.helvetas.org/Publications-PDFs/Switzerland/Jahresbericht-Finanzbericht-Rapport-annuel-financier/EN/JB2023_E_def_web_2.pdf .

HILL, Matthew Alan, 2022. *The rise and fall of democracy promotion in US foreign policy: from Carter to Biden*. New York London: Routledge, Taylor & Francis Group. ISBN 978-1-032-18492-0.

HÖNKE, Jana and LEDERER, Markus, 2013. Development and International Relations. In: *Handbook of International Relations*. Online. SAGE Publications. ISBN 978-1-4739-7119-6. Available from: <http://www.sageereference.com/hdbk_intlrelations/Article_n10.html>.

HSIEH, Hsiu-Fang and SHANNON, Sarah E., 2005. Three Approaches to Qualitative Content Analysis. *Qualitative Health Research*. November 2005. Vol. 15, no. 9, p. 1277–1288. DOI 10.1177/1049732305276687.

HWANG, Hokyu and POWELL, Walter W., 2009. The Rationalization of Charity: The Influences of Professionalism in the Nonprofit Sector. *Administrative Science Quarterly*. June 2009. Vol. 54, no. 2, p. 268–298. DOI 10.2189/asqu.2009.54.2.268.

ISLAM, M. Rezaul, 2016. *NGOs, Social Capital and Community Empowerment in Bangladesh*. 1st ed. 2016. Singapore: Springer Singapore : Imprint: Palgrave Macmillan. ISBN 978-981-10-1747-6.

JAD, Islah, 2007. NGOs: between buzzwords and social movements. *Development in Practice*. August 2007. Vol. 17, no. 4–5, p. 622–629. DOI 10.1080/09614520701469781.

JORDAN, Lisa and VAN TUIJL, Peter (eds.), 2007. *NGO accountability: politics, principles and innovations*. Repr. London: Earthscan. ISBN 978-1-84407-367-2.

KALDOR, Mary, 2003. *Global civil society: an answer to war*. Cambridge, UK : Malden, MA: Polity Press ; Distributed in the USA by Blackwell Pub. ISBN 978-0-7456-2757-1.

KEOHANE, Robert Owen, 2005. *After hegemony: cooperation and discord in the world political economy*. 1st Princeton classic ed. Princeton, N.J: Princeton University Press. A Princeton classic edition. ISBN 978-0-691-12248-9.

KRASNER, Stephen D., 1992. Realism, Imperialism, and Democracy: A Response to Gilbert.

Political Theory. 1992. Vol. 20, no. 1, p. 38–52.

LEWIS, David, KANJI, Nazneen and THEMUDO, Nuno S., 2021. *Non-governmental organizations and development*. Second edition. Abingdon, Oxon New York, NY: Routledge. Routledge perspectives on development. ISBN 978-1-138-35281-0.

LI, Tania Murray, 2007. *The Will to Improve: Governmentality, Development, and the Practice of Politics*. Durham: Duke University Press. ISBN 978-0-8223-8978-1.

MAIER, Florentine, MEYER, Michael and STEINBEREITHNER, Martin, 2016. Nonprofit Organizations Becoming Business-Like: A Systematic Review. *Nonprofit and Voluntary Sector Quarterly*. February 2016. Vol. 45, no. 1, p. 64–86. DOI 10.1177/0899764014561796.

MARTIN, Lisa L. and SIMMONS, Beth A., 2013. International Organizations and Institutions. In: *Handbook of international relations*. 2nd ed. London: Sage. ISBN 978-1-84920-150-6.

MAYRING, Philipp, 2000. Qualitative Content Analysis. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*. 30 June 2000. Vol. Vol 1, p. No 2 (2000): Qualitative Methods in Various Disciplines I: Psychology. DOI 10.17169/FQS-1.2.1089.

MÉDECINS SANS FRONTIÈRES, 2024. *International Financial Report*. Online. Available from: https://www.msf.org/sites/default/files/2024-06/2023%20International%20Financial%20Report.pdf

MIGUEL, Edward and KREMER, Michael, 2004. Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities. *Econometrica*. January 2004. Vol. 72, no. 1, p. 159–217. DOI 10.1111/j.1468-0262.2004.00481.x.

MOL, Annemarie, 2003. *The Body Multiple: Ontology in Medical Practice*. Online. Duke University Press. ISBN 978-0-8223-8415-1. [Accessed 23 November 2024].

MORAVCSIK, Andrew (ed.), 2008. The New Liberalism. In: , *The Oxford handbook of international relations*. Oxford ; New York: Oxford University Press. Oxford handbooks of political science. ISBN 978-0-19-921932-2.

MORTON, Bill, 2013. *AN OVERVIEW OF INTERNATIONAL NGOs IN DEVELOPMENT COOPERATION*. Online. Case Study. UNDP. POSSIBILITIES FOR SOUTH-SOUTH COOPERATION? Available from: https://www.undp.org/sites/g/files/zskgke326/files/migration/cn/UNDP-CH11-An-Overview-of-International-NGOs-in-Development-Cooperation.pdf

MURTAZA, Niaz, 2012. Putting the Lasts First: The Case for Community-Focused and Peer-Managed NGO Accountability Mechanisms. *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*. March 2012. Vol. 23, no. 1, p. 109–125. DOI 10.1007/s11266-011-9181-9.

NÄGELE, Carl Alexander, 2024. Non-Governmental Organizations as Foreign Agents – Foreign Funding of NGOs in Domestic and International Law. . Online. 6 May 2024. DOI 10.18452/28537. [Accessed 19 October 2024].

NEUENDORF, Kimberly A., 2017. *The Content Analysis Guidebook*. Online. 2455 Teller Road, Thousand Oaks California 91320: SAGE Publications, Inc. ISBN 978-1-4129-7947-4. [Accessed 29 November 2024].

NEUMANN, Iver B., 2002. Returning Practice to the Linguistic Turn: The Case of Diplomacy.

Millennium: Journal of International Studies. July 2002. Vol. 31, no. 3, p. 627–651. DOI 10.1177/03058298020310031201.

NICOLINI, Davide, 2009. Articulating Practice through the Interview to the Double. *Management Learning*. April 2009. Vol. 40, no. 2, p. 195–212. DOI 10.1177/1350507608101230.

NICOLINI, Davide, GHERARDI, Silvia and YANOW, Dvora, 2015. *Knowing in organizations: a practice-based approach*. London: Routledge. ISBN 978-1-315-29097-3.

OECD, 1994. *Development Assistance Committee Report 1993*. 1994.

OECD, 2019. *Better Criteria for Better Evaluation: Revised Evaluation Criteria Definitions and Principles for Use*. Online. OECD. ISBN 978-92-64-70943-0. [Accessed 22 November 2024].

OECD, 2023. *Glossary of Key Terms in Evaluation and Results-based Management in Sustainable Development (Second Edition)*. Online. OECD. ISBN 978-92-64-38403-3. [Accessed 22 November 2024].

OECD, 2024. *Official development assistance*. 2024.

ORTNER, Sherry B., 1984. Theory in Anthropology since the Sixties. *Comparative Studies in Society and History*. 1984. Vol. 26, no. 1, p. 126–166.

OSSEWAARDE, Ringo, NIJHOF, André and HEYSE, Liesbet, 2008. Dynamics of NGO legitimacy: how organising betrays core missions of INGOs. *Public Administration and Development*. February 2008. Vol. 28, no. 1, p. 42–53. DOI 10.1002/pad.472.

OXFAM, 2021. *Oxfam International Financial Statements 2019-20.pdf*. Online. Available from: https://oi-files-d8-prod.s3.eu-west-2.amazonaws.com/s3fs-public/2021-11/Oxfam%20International%20Financial%20Statements%202019-20.pdf

PATTON, M. Q. Qualitative Research and Evaluation Methods. 4th ed. Thousand Oaks: SAGE Publications, 2014. ISBN 978-1-4129-7212-3.

PAUL, Samuel, ISRAEL, Arturo, and WELTBANK (eds.), 1991. *Nongovernmental organizations and the World Bank: cooperation for development*. Washington, D.C: World Bank. World Bank regional and sectoral studies. ISBN 978-0-8213-1924-6.

PEOPLE IN NEED, 2024. *People in Need Annual Report 2023*. Online. Available from: https://www.peopleinneed.net/media/publications/2381/file/pin_ar2023_web.pdf .

POTTER, W. James and LEVINE-DONNERSTEIN, Deborah, 1999. Rethinking validity and reliability in content analysis. *Journal of Applied Communication Research*. 1 August 1999. Vol. 27, no. 3, p. 258–284. DOI 10.1080/00909889909365539.

RAHMAN, Mustafizur and FARIN, Sherajum Monira, 2019. *Southern Voice*. Online. 2019. Vol. 53. Available from: https://southernvoice.org/wp-content/uploads/2019/10/Ocassional-Paper-Series-No.53-new.pdf

RECKWITZ, Andreas, 2002. Toward a Theory of Social Practices: A Development in Culturalist Theorizing. *European Journal of Social Theory*. May 2002. Vol. 5, no. 2, p. 243–263. DOI 10.1177/13684310222225432.

REUS-SMIT, Christian and SNIDAL, Duncan (eds.), 2008. *The Oxford handbook of international*

relations. Oxford ; New York: Oxford University Press. Oxford handbooks of political science. ISBN 978-0-19-921932-2.

RISSE, Thomas, 2013. Transnational Actors and World Politics. In: CARLSNAES, Walter, RISSE, Thomas and SIMMONS, Beth, *Handbook of International Relations*. Online. 1 Oliver's Yard, 55 City Road, London EC1Y 1SP United Kingdom: SAGE Publications Ltd. p. 256–274. ISBN 978-0-7619-6305-9. [Accessed 10 October 2024].

ROBINSON, Mark, 1994. Governance, democracy and conditionality: NGOs and the New Policy Agenda. In: CLAYTON, Andrew (ed.), *Governance, democracy, and conditionality: what role for NGOs?*. Oxford: INTRAC. INTRAC NGO management & policy series, 2. ISBN 978-1-897748-01-5.

ROUSE, Joseph, 1994. *Knowledge and power: toward a political philosophy of science*. 1. publ. Cornell paperbacks, 2. print. Ithaca, NY: Cornell Univ. Press. Cornell paperbacks. ISBN 978-0-8014-1959-1.

ROY, Ananya, 2010. *Poverty capital: microfinance and the making of development*. First edition. New York: Routledge. ISBN 978-0-203-85471-6.

ROY, Arundhati, 2014. The NGO-ization of Resistance. *Toward Freedom*. 2014. Vol. 8, no. 2.

SACHS, Jeffrey D, 2012. From Millennium Development Goals to Sustainable Development Goals. *The Lancet*. June 2012. Vol. 379, no. 9832, p. 2206–2211. DOI 10.1016/S0140-6736(12)60685-0.

SADOUN, Britta, 2006. Donor Policies and the Financial Autonomy of Development NGOs. *Development*. June 2006. Vol. 49, no. 2, p. 45–51. DOI 10.1057/palgrave.development.1100256.

SAVE THE CHILDREN INTERNATIONAL., 2023. *Trustees' Report and Financial Statements for 2022.*. Online. Available from: https://resourcecentre.savethechildren.net/pdf/SCI-Trustees-Report-and-Financial-Statements-2022.pdf/

SCHATZKI, Theodore R. (ed.), 2002. *The site of the social: a philosophical account of the constitution of social life and change*. University Park: Pennsylvania State University Press. ISBN 978-0-271-02292-5.

SCHREIER, Margrit, 2012. *Qualitative Content Analysis in Practice*. Online. 1 Oliver's Yard, 55 City Road London EC1Y 1SP: SAGE Publications Ltd. ISBN 978-1-84920-593-1. [Accessed 29 November 2024].

SIMMONS, P. J., 1998. Learning to Live with NGOs. *Foreign Policy*. 1998. No. 112, p. 82. DOI 10.2307/1149037.

SMILLIE, Ian, HELMICH, Henny, and ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (eds.), 1993. *Non-governmental organisations and governments: stakeholders for development*. Paris : [Washington, D.C: Development Centre of the Organisation for Economic Co-operation and Development ; OECD Publications and Information Centre, distributor]. ISBN 978-92-64-03899-8.

SPIEGEL, Gabrielle M., 2005. *Practicing history: new directions in historical writing after the linguistic turn*. New York: Routledge. ISBN 978-0-203-33569-7.

THEMNÉR, Lotta and WALLENSTEEN, Peter, 2012. Armed Conflicts, 1946–2011. *Journal of*

Peace Research. July 2012. Vol. 49, no. 4, p. 565–575. DOI 10.1177/0022343312452421.

THOMPSON, Simon, 2010. Models of Democracy and the Politics of Recognition. In: SEYMOUR, Michel (ed.), *The Plural States of Recognition*. Online. London: Palgrave Macmillan UK. p. 122–140. ISBN 978-1-349-31228-3. [Accessed 15 October 2024].

TICKNER, Judith Ann, 2014. *A feminist voyage through international relations*. New York: Oxford University Press. Oxford studies in gender and international relations. ISBN 978-0-19-937471-7.

UN (ed.), 2003. *Handbook on non-profit institutions in the system of national accounts*. New York: United Nations. Studies in methods, no. 91. ISBN 978-92-1-161461-9.

UN, 2015. Transforming our world: the 2030 Agenda for Sustainable Development | Department of Economic and Social Affairs. . Online. 2015. Available from: https://sdgs.un.org/2030agenda [Accessed 12 November 2024].

VEDDER, Anton (ed.), 2007. *NGO involvement in international governance and policy: sources of legitimacy*. Leiden Boston: Martinus Nijhoff Publishers. Nijhoff law specials, v. 72. ISBN 978-90-474-2243-3.

VIVIAN, Jessica, 1994. NGOs and Sustainable Development in Zimbabwe: No Magic Bullets. *Development and Change*. January 1994. Vol. 25, no. 1, p. 167–193. DOI 10.1111/j.1467-7660.1994.tb00513.x.

WALLER, Michael, 1993. *The end of the communist power monopoly*. Manchester New York: Manchester university press. ISBN 978-0-7190-3819-8.

WALTZ, Kenneth (ed.), 1986. Political Structures. In: , *Neorealism and its critics*. 7. print. New York: Columbia Univ. Press. The political economy of international change. ISBN 978-0-231-06349-4.

WENDT, Alexander (ed.), 1999. *Social theory of international politics*. Cambridge, U.K New York: Cambridge University Press. Cambridge studies in international relations, 67. ISBN 978-0-511-61218-3.

WHITE, Howard, 2005. *Challanges in evaluating development effectiveness*. Brighton: Institute of Development Studies. Working paper / Institute of Development Studies, 242. ISBN 978-1-85864-854-5.

WHITE, Howard, 2014. Current Challenges in Impact Evaluation. *The European Journal of Development Research*. January 2014. Vol. 26, no. 1, p. 18–30. DOI 10.1057/ejdr.2013.45.

WOHLFORTH, William C. (ed.), 2008. Realism. In: , *The Oxford handbook of international relations*. Oxford ; New York: Oxford University Press. Oxford handbooks of political science. ISBN 978-0-19-921932-2.

WOODS, Adéle, 2000. *Facts about European NGOs Active in International Development*. Paris: Organization for Economic Cooperation & Development. Development Centre Seminars. ISBN 978-92-64-18548-7.

WORLD VISION INTERNATIONAL, 2023. *2022 Consolidated Financial Statements.*. Online. Available from: https://www.wvi.org/publications/financial-statements/consolidated-financial-statements-2022-and-2023

YIN, R. K. Case Study Research and Applications: Design and Methods. 6th ed. Los Angeles: SAGE Publications, 2018. ISBN 978-1-5063-3616-9.

ZEHFUSS, Maja, 2013. Critical Theory, Poststructuralism, and Postcolonialism. In: *Handbook of International Relations*. Online. SAGE Publications. ISBN 978-1-4739-7119-6. Available from: <http://www.sageereference.com/hdbk_intlrelations/Article_n10.html>.

Abbreviations

3ie — International Initiative for Impact Evaluation

CSR — Corporate Social Responsibility

DAC — Development Assistance Committee

DCA — Directed Qualitative Content Analysis

EC/EU — European Commission / European Union

ECOSOC — Economic and Social Council

GONGO — Government-Organised Non-Governmental Organisation

INGO — International Non-Governmental Organisation

IO — International Organisation

IR — International Relations

MDG — Millennium Development Goal

MNC — Multinational Corporation

MSF — Médecins Sans Frontières (*Doctors Without Borders)*

NGO — Non-Governmental Organisation

NNGO — Northern Non-Governmental Organisation

NPA — New Policy Agenda

NPM — New Public Management

ODA — Official Development Assistance

OECD — Organisation for Economic Co-operation and Development

PM — Project Manager

RCT — Randomized Controlled Trial

RQ — Research Question

SDG — Sustainable Development Goal

SNGO — Southern Non-Governmental Organisation

UN — United Nations

List of Appendices

Appendix no. 1: Overview of the number and type of analysed primary resources (Table)

NGO	Interview	Strategy	Annual Report	Learning Document/Brief	Evaluation
Concern	1	1	2	7	2
Helvetas	1	1	2	10	5
People in Need	2	3	2	4	1