OSLOMET

Maria M. Bakken

How Can We Learn from Evaluations?

A Qualitative Study on Learning from Evaluations in Norwegian Development Cooperation

Master´s thesis in International Social Welfare and Health Policy

Oslo Metropolitan University, Faculty of Social Science

May 2023

Acknowledgments

OSLOMET

Abstract

This thesis seeks a fundamental understanding of the connection between evaluations and learning in Norwegian development cooperation through four objectives shaped by a literature review. The objectives are to identify and analyze; 1. Attitudes towards the usefulness of evaluations of development cooperation. 2. How methodological considerations in evaluations can affect learning. 3. How evaluations are believed to be necessary for changes within Norad. 4. How attitudes towards evaluations may be affected by organizational factors. Previous research points out interesting factors that inhibit learning from evaluations from different angles. This thesis adds perspectives from inside development cooperation to explain their professional learning experiences. The analysis and discussion utilize a synthetic theoretical view based on evaluation methods (Weiss) and organizational learning loops (Argyris & Schön) because the views supplement each other in coverage of the research objectives. Future research on the topic would benefit from including theoretical perspectives that further explain the dynamics between research and policymaking in the theoretical perspective.

Data collection consists of eight semi-structured interviews, with different staff in Norad and one researcher. The thematic analysis shows that learning from evaluations is connected to four overarching themes. The themes concern; 1. Attitudes Towards Evaluation Methods. 2. Organizational Factors. 3. Political Influences. 4. Evaluation Utilization.

The findings show that professionals care about evaluation quality; if they find this lacking, they will be less open to learning. It is therefore essential to build cogent research to create organizational learning where staff can trust its findings. Evaluation research is impacted by degrees of objectivity and the involvement of external consultants, where negative experiences of this create less openness for learning. Dissemination strategies may be necessary for both learning and accountability, and this is dependent on individual practices rather than incorporated structures. Currently, there is no structure to support learning, and professionals find this problematic in practice. The findings present situations that primarily support single-loop learning from evaluations. Development cooperation should support further learning culturally and systematically, and such shifts must be prioritized in practice.

List of Abbreviations

Norad   Norwegian Agency for Development Cooperation

MFA   The Ministry of Foreign Affairs

MCE   The Ministry of Climate and Environment

Table of Contents

CHAPTER ONE

INTRODUCTION

"Evaluation, at its best, distinguishes *what works from what doesn't,* and it helps separate effective change makers from resource wasters, boastful charlatans, incompetent meddlers, and corrupt self-servers." (Patton, 2008, p. xviii)

1.1 Presentation of the Research Problem

From a global perspective, development aid has been subject to much research and discussion through several decades of experience. Despite this, there is still a significant lack of knowledge of what increases positive outcomes in recipient countries, and what may do the opposite (Gaarder & Bartsch, 2015). One cannot, from a global perspective, determine whether or not official development assistance is doing more good than harm. Because the answer to this depends on many external factors, such as power relations between groups and governance structure (World Bank Group, 2017, p. 26). Evaluations, therefore, constitute a defining feature of development cooperation, which is needed to learn of the effects of development aid. Evaluations are central to a shared understanding of important international social and economic questions. The following literature review suggests that evaluations serve a variety of functions in the Norwegian Agency for Development Cooperation (Norad). Yet, some argue that evaluations are not used to their full extent as guiding principles for organizational adjustments. It is suggested that it is not necessary to focus on heavier evaluation, but that such evaluations should be used more strategically in Norwegian development cooperation (Tjønneland, 2018).

Debates on evaluations of Norwegian development cooperation point to a need of greater understanding of professional learning experiences and how such evaluations should be utilized. A recent literature review points out that despite the growth in numbers of evaluations, this does not mean knowledge and learning is following, but may create an overload of information that is difficult to navigate (Reinertsen et al., 2022). And it has been suggested that the structure of the Norwegian aid system is not focused towards results, which means they can prioritize project funding over quality evaluations and knowledge

1

building (Lindkvist & Dixon, 2014). This is also found on the individual level where evaluations score low amongst learning arenas for staff in the aid administration because other arenas may be more important (Samset, 2015; Samset et al., 1993, p. 64). By analyzing what professionals currently or previously employed in Norad think about evaluations as learning opportunities, one could better understand necessary evaluative or organizational considerations for learning. Moreover, this has the potential for future evaluations to consider professional experiences when directing evaluation designs.

The objectives of the study are as follows:

1. *To identify and analyze attitudes towards the usefulness of evaluations of development cooperation*

2. *To identify and analyze how methodological considerations in evaluations can affect learning*

3. *To identify and analyze how evaluations are believed to be necessary for changes within Norad*

4. *To identify and analyze how attitudes towards evaluations may be affected by organizational factors*

The overall research question for this thesis is:

What are the professional attitudes towards learning from evaluations within the Norwegian Agency for Development Cooperation, and how may this be understood in an organizational perspective?

This introduction elaborates the research question in the following way: First, presenting what this study may add to existing knowledge through its current relevance. Second, providing a contextual overview of the Norwegian development cooperation system. Third, a description of the literature review with inclusion criteria identifying relevant literature. And lastly, showing what is known from previous research about perceptions of learning from evaluations.

1.2 Relevance of the Study

This thesis situates itself amongst significant questions of learning about the effects of development programs and making decisions based on that knowledge. As the literature review will show, many scholars have researched this topic from the perspective of staff in development cooperation, specific development projects, general procedures and methods in Norad, evaluation quality, or learning. This thesis explicitly narrows in on an in-depth focus on professionals within the field of Norwegian development cooperation, to explain their experiences of learning through an organizational perspective on Norad. Previous research provides information about many factors that may inhibit learning within Norad. This thesis seeks to contribute to the current state of knowledge through in-depth semi-structured interviews. This can help provide answers to the complex topic of organizational learning from evaluations by considering how staff in Norad relate to this. Through using theoretical perspectives from the field of evaluation methods and organizational learning, this thesis will investigate the link between such considerations. It may be essential to be able to provide for and practice organizational learning through evaluations to conduct the best development cooperation possible. What follows in the chapters for results and discussion of this thesis is based on the personal thoughts of eight anonymous participants on this topic, providing answers to attitudes of learning from evaluations.

1.3 Brief Overview of the Norwegian System for Development Cooperation

Norway is one of the most generous donors of official development assistance in the Development Assistance Committee in the Organization for Economic Co-operation and Development (OECD). With commitments to use one percent of the gross national income (GNI) on official development assistance, Norway has been in the top five of the most generous donor countries since 1976 (OECD, 2022). In 2022, the Norwegian state budget proposal projected spending 0,75 percent of GNI on official development assistance, which is less than previous years. However, the budget argued for an increase in expected GNI, meaning that this still constitutes an increase in the development budget from last year (Prop. 1 S Gul bok (2022–2023), pp. 17-18). On the basis of being a generous donor committed to development, this thesis seeks an essential understanding of how knowledge of outcomes are gained and transferred within the system and how the staff perceives the functionality of this.

The official development aid from Norway is produced through governmental propositions. The propositions include state budgets that allocate funding to the development sector. The thematic areas and goals for development cooperation are also set through governmental proceedings. Based on the outcome of this, the Ministry of Foreign Affairs (MFA) will form strategies for cooperation with selected countries. The MFA is politically responsible for most of the official foreign development cooperation, concerning state-to-state partnerships and multilateral cooperation. Norad is subject to the MFA, and the Ministry of Climate and Environment (MCE) with regards to selected thematic areas. Their main tasks is to contribute to effectively managing funds for development, quality assurance of development cooperation, and evaluating and communicating results of such work. Their funding is managed through voluntary and international organizations, and through research on development (Norad, n.d.-b). As stated in a recent yearly report, Norad is to make sure that management of development aid is systematically based on knowledge. Their tasks include managing grants in the development field where they in 2021 managed about half of the Norwegian aid budget. They also state knowledge-sharing and professional advice towards different authorities as an important part of their responsibility (Norad, 2022b, pp. 27-32).

Norad´s strategy towards 2030 underlines strengthening their expertise through the exchange of knowledge. Aiming for goals through holistic and strategic approaches and integrated use of knowledge in all work areas. In innovative matter it states that research should always be applied to gain insight and understanding, and that this is more important than established routines. This happens under the current slogan stating that "Facts Inform Policy" (Norad, 2021). This ambitious goal creates a need to understand how this is experienced in practice by the staff in Norad, and how evaluations may play a role in such learning.

The evaluation department of Norad initiates and conducts evaluations, independent of those branches responsible for managing development programs. This department is predominantly responsible for overseeing and documenting the effectiveness of given aid and if relevant goals are met. At the same time, they list one of their primary purposes as being contributors to learning and keeping different professionals within the foreign aid field

accountable. In their latest yearly report, they state that evaluations should contribute to openness by providing information to the public and promoting findings that are useful in creating budgets and further development of the function being evaluated (Norad, 2022a). The purpose to promote learning through evaluation findings that promote further development, underlines learning as a central goal in Norad. This thesis seeks knowledge of the experiences of this achievement and how this is facilitated through evaluations.

Norad distinguishes between centralized and decentralized evaluations. Some areas of development cooperation is evaluated centrally by the department for evaluation, whereas most evaluations are conducted decentralized by other authorities (Norad, n.d.-a). Decentralized evaluations of programs and projects are conducted based on calls from the MFA, Norad, or different Norwegian embassies. Decentralized evaluations can serve as necessary knowledge with regards to the effect of development aid, while not being part of the independent evaluations conducted by the department itself (Norad, 2022a, p. 16). Centralized evaluations under the department for evaluation will go through a typical process where the department is responsible for choosing what to evaluate and what questions to answer, creating a "term of reference" together with chosen experts. The evaluation mission will then be sent out through international tenders, to hire consultants who will conduct the evaluation independent of the evaluation department. The department will later receive drafts of the report to consider its quality. The evaluation department will then share the final work with all involved in the program, allowing stakeholders to give feedback. Based on this, the evaluation department will consider recommendations on what needs to be followed up in Norwegian development politics, while also publishing and presenting the evaluation report. The MFA, or the MCE, is in charge of creating a plan as to how to manage the response. The evaluation department will incorporate this into their yearly published report to summarize the results of this year's evaluations and the use of resources (Norad, n.d.-a).

The research question and objectives are operational to study the Norwegian system due to considering different methods of conducting evaluations which includes looking into both centralized and decentralized evaluations, to understand how professionals relate to this. Considering how staff relate to evaluations in their work and how evaluations is seen as

necessary for changes will give insight into how the organization fulfills their goals. Finally, the thesis accounts for organizational factors that may affect such attitudes and actions. An in-depth perspective from inside the Norwegian system for development cooperation is meaningful to understand how they strive to fulfil the mandate for learning and knowledge utilization, while recognizing divergent evaluation methods and impactful organizational features.

Literature Review

1.4 Literature Search Strategy

The literature search for this thesis was conducted throughout the course SIW4500 as part of the Literature Review Project, and the presentation builds on previous work conducted in the course (Bakken, 2022). The aim of the review was to create a thorough introduction to the field of evaluations within Norwegian development cooperation. This literature review was meant to serve as a background for the thesis, providing a guide for decisions to be made in terms of research design. Additionally, it intended to provide an overview of the current state of knowledge through research and identify gaps which this study intends to provide perspectives on (Xiao & Watson, 2019).

The literature search occurred between September 2022 until December 2022, which included large amounts of literature to gauge based on initially open inclusion criteria. The inclusion criteria were to find studies conducted on Norway, or that had relevancy to Norwegian development aid. Topics were also to surround evaluation usage or methods for evaluating international development aid. Due to an initially broad topic, the search had to include wide-ranging criteria and many databases to gain an overview of the field. However, this resulted in relatively high sensitivity, due to the attempt to maximize the number of potentially relevant hits. This also resulted in a substantial amount of non-relevant literature that had to be removed manually. The ratio between sensitivity and specificity was not optimal, but due to the low number of relevant hits at the start of the search, it was advised by a librarian to widen the search and instead remove what is found to be irrelevant (Booth, 2016). A librarian also advised to conduct additional searches, since many known and relevant articles did not show through the database search. Therefore, a hand search in the journal "Stat & Styring" was added, dating five years back, and a site search in "Journal of

Development Effectiveness". Additionally, I reached out to two experts in the field to ask for guidance on relevant literature and conducted a small citation search of central articles. See the PRISMA Flow Chart and explanation in Appendix 1 and 2 for a thorough overview of the literature search. The literature will be presented to give insight into what constitutes difficulties with learning from and using evaluations, as well as the perceived causes of this.

1.5 The Quality and Utilization of Decentralized Evaluations

An evaluation of the quality of decentralized evaluations in Norwegian development cooperation found that usage of such evaluations is mostly done instrumentally. This means that knowledge from evaluations are directly used to inform an ongoing policy or program, mainly by those responsible for grant management. The symbolic use of reviews has also been shown to be relatively high, as a means to justify decisions or routines before ending an intervention. Additionally, conceptual use was somewhat limited, classifying how knowledge may change how professionals think about what works and what does not (Chapman et al., 2017). Carol Weiss originally defines the classification of different ways to use evaluations. According to her, evaluations are used instrumentally when results help policymakers fathom decisions, and conceptually when the findings change perceptions about the program's purposes (Weiss, 1998b). The classifications are connected to her investigation into how research is used in policymaking presenting seven different views on research utilization through reviewing literature in the field (Weiss, 1979).

The evaluation of decentralized evaluations, constituting a meta-evaluation, also pointed out how decisions based on such decentralized evaluations of programs were often based on findings and recommendations through unreliable methods. The authors found that over 65 percent of decentralized evaluations of development programs did not include sound methods and analysis, based on a limited or non-existent discussion on methodological work. The authors found that decentralized evaluations scored low on quality criteria for methods, such as analysis, ethics, and data collection, which weakens the chance of producing credible findings (Chapman et al., 2017). Two follow-up studies have been conducted on this topic since 2017, in an attempt to improve the quality of decentralized evaluations. Findings from 2020 show that the quality was still low, where approximately half of the 55 evaluations under

review had less than adequate quality based on half of the selected quality criteria. Methodology was found as the weakest quality criteria and the reports rarely described analytical frameworks. The study also found infrequent presentation of selection and samplings methods, and lack of critical assessment of data (Ternström et al., 2020). Results from 2021 show similar conclusions, that while evaluation reports responds to the purpose of the evaluation, they do not as often explain how they came to their conclusions. Which limits both credibility and learning through evaluations. Additionally, excluding contextual information in evaluations limits usefulness of evaluations for wide range and future staff (Ternström et al., 2021).

1.6 Organizational Factors in Norwegian Development Cooperation

There are some organizational factors that may impact knowledge and learning amongst the staff in Norad. As research shows, the capacity to learn specific information, such as in a technical field, is rather high. However, absorbing more abstract knowledge regarding social and political affairs is not as common. This is explained through competence needs, high organizational diversity which prevents shared masses of knowledge, and no existence of a dominating structure for knowledge. The more knowledge and experience an organization has, the more advanced its structures for knowledge will be. This makes it easier to learn and change once information is incorporated into advanced systems, with more sources to gather information. The system in place when this study was conducted indicated that the organization needed to question their foundational assumptions and change based on this. This change would help in achieving grander learning opportunities. It was also found that learning by facilitating involvement and communication in the evaluation process was important, suggesting that staff take an active part in the evaluation (Forss et al., 1994). Furthermore, the aid budget in Norway has increased in the last 15 years, which may constitute a factor for decreased competence of professionals within development cooperation (Tjønneland, 2018).

Also constituting organizational challenges of learning may be the renewal of personnel within the field of development cooperation. Renewing personnel working on long-term projects may cause challenges for using evaluations as a constant learning possibility, because

it does not become a force in decision-making. The achievement of set goals often relies on being evaluated in ex-post ratings conducted after the program is finished. Those originally responsible for quality and achievement of goals may have moved on in their career at the point of this evaluation. This means that the ex-post evaluation will no longer serve as an incentive in their work or something that will hold them accountable. The staff shift is a significant problem to overcome, as incentives could translate to professional and organizational learning, where such evaluations could be a rich source of information for all professionals. To tackle this time inconsistency, it is proposed to assess evaluability at the start of programs when professionals will be held accountable for their current projects (Gaarder & Bartsch, 2014).

1.7 Ensuring Evaluability

Norwegian development cooperation does not impose standards for reporting results, leaving many decisions regarding interventions within the discretion of the partner country. Norwegian political leadership states the need for results measurement firmly. However, many professionals find that this does not equal a drive for a focus on results through senior management. There has also been stipulated a shared feeling of failure to recognize implications on resources of measuring results (Lloyd et al., 2014). This creates complications for informed decisions of professionals to judge the quality of the recipient's approach and ensure evaluability. It has been suggested that the Norwegian aid system is not developed to provide documentation to measure results and that the staff is not trained with enough competence. Furthermore, the limited use of quality assurance functions does not promote evaluability, where components that support evaluability are not used appropriately in practice (Villanger et al., 2016). Ensuring evaluability and learning may be a complicated task in today's aid environment, due to its growing ambitions in many areas for development. The ever-increasing aims create difficulties in establishing links of causes and effects in different societies. Which again makes it problematic to identify impact evidence for a specific program. Methodological challenges for evaluation include separating impact from aid in relation to other factors that may have influenced observed outcomes (Villanger & Jerve, 2009).

An evaluation looking into why grants supported by the Norwegian aid administration prove difficult to evaluate, found that grants often do not have adequate frameworks for results. This consequently makes it difficult to judge the larger successfulness of Norwegian development cooperation. Staff report that time pressure and low priority of senior management reduce their effectiveness in reviewing and monitoring grant performance. Reporting at the end of programs is especially found to be weak which reflects a lack of guidance and planning for evaluations. An assessment of evaluation reports found that the reports are suitable to provide information of outputs, but not causality (Norad, 2014).

1.8 Trade-offs for Learning

The primary purpose of conducting and using evaluations and their intended purpose is a contested debate. Some scholars argue that the primary utility of evaluations are to provide individual and organizational learning (Forss et al., 1994). However, seeing learning as the primary goal of evaluations may not be the general view within professional perspectives on the topic. In contrast, a study conducted in 1993 shows that professionals within the Norwegian aid administration would score evaluations at place 15 out of 19 in total in a range of the importance of different arenas for learning within aid. This low score for evaluations may be because humans learn more through engagement than by reading long evaluation reports. In comparison, the more critical arenas for learning were contact with other organizations, local inspections, and the exchanging of information between professionals (Samset, 2015; Samset et al., 1993, p. 64).

To understand how evaluations may be helpful for learning, one must see what is possibly stopping this from being the main priority. Studies presents the dual perspective on purposes of evaluations of aid in Norad between accountability and learning. Building on previous studies on the relationship between learning and accountability, a literature review found accountability in three different forms. They include accountability towards donors, recipients, and other actors involved in the project. It is theorized that learning can be joined with accountability concerning other actors involved in the project. The two different purposes of evaluations used in Norad, concerning prioritized orders and diverging methodologies, may not be irreconcilable. Many believe there to be a trade-off between

prioritizing accountability for measures taken and learning within evaluations. Nonetheless, they are possible to combine, and learning may not need to be sacrificed based on the demand for accountability. However, to achieve learning, professionals need to recognize the necessary trade-offs between accountability and learning. As long as accountability remains the main focus of evaluations of Norwegian foreign aid, large-scale learning will constitute an illusion (Reinertsen et al., 2022).

Perspectives that combine learning and accountability from evaluations is contested. In comparison, the possible irreconcilability between accountability and learning may explain the trend to emphasize monitoring programs at the expense of evaluations. The differences in methodology, incentives, and data, may create models for evaluation that do not facilitate learning. Development aid is guided by managing based on results which emphasize monitoring impact and evaluating effectiveness. This focus may take away space from impact evaluation and lessons learned through reducing non-linear human development and change, to logic frameworks predicting outcomes from interventions (Armytage, 2011).

1.9 Time and Recourses for Learning

Literature emphasize that continuity is essential in seeing results from development programs. This is related to tendencies for rapid turnovers towards new ideas when faced with inadequate outcomes (Eidhammer, 2007). This aspect may also relate to organizational factors and incentives. A study focusing on the motivating factors of giving aid in Norway found that learning from successful or unsuccessful programs is necessary to prioritize good results in the welfare of recipients. Avoiding this constitutes a situation where one can constantly implement new projects, rather than funding fewer projects and investing in solid evaluations. This means that we will have less knowledge on what should be scaled up or down in terms of benefitting recipients (Lindkvist & Dixon, 2014).

1.11 Outline of the Thesis

To systematically provide a comprehensive answer to the research question and objectives, the thesis is divided into 6 chapters including this introduction. Chapter two explains the synthetic use of two theoretical frameworks from separate fields, that will lead the way for

understanding the relationship between evaluations and organizational learning. Chapter three explains the methods applied to gather and analyze data, as well as reasons for why such methods were chosen. Chapter four presents the empirical findings from the research through four themes that relate to the research question and objectives. The themes present the different aspects of learning and evaluations that were most present in the interviews and aligns simultaneously across the different objectives and research question. Chapter five is used to discuss the essence of the answers to the research question and objectives, considering previous research and the relevance of the theoretical perspective. The discussion follows the same themes presented in chapter four, where the four objectives and the research question will relate across each theme. Chapter six will conclude the thesis and provide recommendations for future research, and lastly discuss the limitations of the thesis.

CHAPTER TWO

THEORETICAL FRAMEWORKS

The analysis and discussion will be based on parts from two separate theoretical spheres to provide comprehensive reflections and answers to the research question. The topic for this thesis is based on both learning and evaluations, which is why a synthetic use of two theories is believed to provide answers that cover the research topic. Following the rationale for synthesis presented by Cairney (2013), combining insight from different theoretical approaches aims to produce a single suitable theoretical perspective for a specific research topic. The two views in the synthetic perspective are evaluation methods presented by Carol Weiss and organizational learning loops presented by Chris Argyris & Donald Schön. These perspectives are found to be the most relevant for analyzing how methods for conducting and disseminating evaluations may be part of Norad´s learning processes. I propose a point of view where methodological considerations in evaluations may be part of organizational learning in Norad through how evaluations may facilitate different levels of learning when faced with outcomes of programs. The synthetic connection is therefore based on the possible dependence learning has on evaluation strategies. This connection may be important in understanding how people relate to learning, through considering how the evaluations are

produced. Additionally, the synthetic approach will show the way for an organizational perspective in the thesis and provide depth to the research question and objectives.

2.1 Evaluation Theory

There are several traditions for analyzing evaluations and research utilization that would have been suitable to apply in this thesis. Research conducted by Patton (2008), Caplan (1979), and Fixsen et al.(2009), makes great contributions to understanding how research is utilized in decision-making processes. Patton developed the utilization focused evaluation describing methods for conducting evaluations with its intended usage in mind. As the name describes, the utilization focused evaluation premises that evaluations should be valued based on their utility in the real world (Patton, 2008, p. 37). The two communities theory by Caplan categorizes the distinction between researchers and policymakers, with the challenges this creates in utilizing research findings (Caplan, 1979). Fixsen et al. shows interest in the receiver side of programs, through theorizing implementation components of successful welfare programs, suggesting a missing link between science to service through implementation (Fixsen et al., 2009). The thesis leaves out the abovementioned theoretical perspectives, to prioritize evaluation theory found as more relevant and suitable for the topic.

The theoretical work of Carol Weiss will be the primary approach towards evaluations in this thesis, due to its suitable coverage of the topic. Weiss has had tremendous influence in the evaluation field since the 1960s. Her academical legacy include the recognition of political aspects of evaluations, widening the definition of how evaluations can be utilized, and developing a theory-based understanding of the possible significant effects of evaluation. Weiss first theorized that the role of evaluator was consequently political, due to objects under evaluation being the output of political activity, and the possible political effects of the findings. All of which is now seen in conventional evaluation theory, deeming her the "founding mother" of evaluation (Mathison, 2005, p. 449).

2.2 In-house versus Outside Evaluation

As in all other types of research, the methods under which it has been conducted is important for how it is received. This relates to whether evaluations should be conducted by

the organization itself, or by someone external. Naturally, it is essential in evaluation perspectives to recognize the minimization of subjectivity in the process and to overcome perceptive limitations (Lopez-Lee, 1982). Some argue for the ideal self-evaluating organization that constantly monitor their own activities to determine goal achievement (Wildavsky, 1972). While other typical arguments against external evaluations claim that evaluators would only spend time learning parts of a larger project context that someone on the inside would already know. However, independence and externality is important to secure demands for accountability in order to receive accurate pictures and make sound decisions (Lopez-Lee, 1982).

Evaluations in the Norwegian development system are conducted either centralized or decentralized. Centralized evaluations are administered by the department for evaluation where the in-house staff will have much responsibility in shaping the evaluation, while it is conducted by outside consultants hired through international tenders. Decentralized evaluations are initiated by those responsible for measures to be evaluated, and they will leave the evaluation with independent experts (Norad, n.d.-a). The in-house versus outside perspective is important in understanding methodological nuances that can impact the later learning opportunities from both types of evaluation. To analyze the research question and objectives that connects evaluations and learning, it is necessary to understand the underlying mechanism at play when deciding level of interference of inside staff. Weiss (1998a, pp. 37-39) presents this debate of in-house or outside evaluations by considering five factors:

1) Administrative confidence regards relational features and how central administrators need confidence in the professional skills of the evaluators. External evaluators may exist too far from the realities under evaluation and may therefore produce findings without practical value. The competence of the evaluators are therefore key to ensure administrative confidence.

2) Objectivity concern the biases of the evaluators. Comprising a typical argument for external evaluations, through the requirement of removing possibilities to collect and interpret data with a specific angle in mind.

3) Understanding the program that is under evaluation is essential for evaluators. This knowledge may be easier to accumulate by in-house staff who experience its administrative

issues and the real-life events of the program. External evaluators may learn this. However, requiring extra effort and access to information sources.

4) Utilizing evaluation results may require that the evaluators themselves take an active role in pushing findings towards policymakers. In-house staff may move the results through meetings or conferences, increasing the chances of findings being taken seriously. On the contrary, external evaluators may bring prestige and authority needed to attain attention around the results. They also carry the chance to skip hierarchies to bring findings to top management.

5) External evaluators may have greater autonomy of their own work, not accounting for assumptions, organizational arrangements, and existing frameworks. The external evaluators may therefore provide a broader perspective of the program, perhaps drawing their results towards structural changes needed. Changes need to be accounted for in a long time perspective, and a broad perspective may influence habits through fresh ideas reinforced by influential actors. It is important to raise questions that may be uncomfortable for internal professionals abiding by the organization's current conventions, based on the possibility of providing recommendations for their structure (Weiss, 1998a, pp. 37-39).

2.3 Dissemination and Utilization of Evaluations

Evaluation research is generally concerned with how findings from such studies are utilized. Utilization can be seen as evidently that in the absence of the evaluation findings, decision-makers would have acted or thought differently (Leviton & Hughes, 1981). Utilization is consequently connected to learning in Norad, based on the mandate to produce evaluations that are valued based on learning practices. Efforts in dissemination and utilization may therefore impact the chances professionals have to relate to evaluations in their work, and it may affect their beliefs about its usefulness. Dissemination and utilization theory is connected to the field of research and development through the analysis of research utilization in policymaking. Where cooperation with knowledge brokers to present research evidence to policymakers is seen as useful in making social science applicable in government agencies (Lynn, 1978, p. 4). However, dissemination of evaluations may be a diffuse process because this is often not the specific responsibility of anyone in particular, and there is typically a lack of channels for regular dissemination (Weiss, 1978, p. 57). This thesis sees dissemination

strategies as distinct factors within an organizational perspective, in terms of seeking an understanding of how this happens and how it affects learning.

Learning the lessons provided by evaluations may be dependent on the attention of key staff through dissemination. It may be impractical to rely only on the report for learning, because only a few dedicated people will read this, and most will learn the findings through communication. Weiss therefore proposes to facilitate communicative learning from evaluations which suggests including knowledge-users in the research process. This inclusion can facilitate ownership of the evaluation and can be done in decision processes, through giving feedback, and participation in analysis (Weiss, 1998a, pp. 305-307).

Despite dissemination efforts, evaluations may be under-utilized, which is explained through organizational characteristics or political constraint. Characteristics of an organization may define a certain comfortable status quo. Moving away from this may require large resources that are often restricted. This may include re-training and new hires within the staff, which is why new practices will need to provide evidence of significant effect in order to disrupt the existing system. Political constraints define how evaluations are more than scientific procedures, but they also flow into political decisions. Programs under evaluation often come to life through political affairs. After the evaluation, they may become embedded in political decision-making. Different roles also define different interests, where politicians and legislators may seek reelection through satisfying voters. This can affect their tolerance towards research findings (Weiss, 1998a, pp. 310-314).

Evaluations are often disregarded because they disclose inadequacies. To counteract this, it is important to make the evaluation more cogent, because users care abouts its methodological quality as the basis of decisions. Meta-analysis, which synthesizes results across several evaluations showing the total weight of evidence, can create more willingness to apply the findings (Weiss, 1998a, pp. 314-317). Weiss defines meta-analysis as "The systematic analysis of the results of a body of evaluations of similar programs to produce an estimate of overall program effect" (Weiss, 1998a, p. 333). This term is often used for meta-evaluations, and will be understood as such in later discussions (Weiss, 1998a, p. 236). Meta-

analysis may create greater confidence in the results and heighten the motivation to utilize evaluations. It may also be necessary to increase efforts in dissemination. To provide results that will be directly utilized in political affairs, it is also recommended to report on such evaluations with awareness of the political atmosphere. Lastly, one cannot expect changes to happen rapidly. The concept of utilization of evaluations is not a mechanical process, but something that entangles other factors, such as experience, political processes, and ideological commitment (Weiss, 1998a, pp. 314-317).

2.4 Theory on Organizational Learning

Learning may be seen as both cognitive and behavioral change. Cognition emphasizes changes in knowledge based on experience and behavioral learning revolves around changes in one´s behavior. The two learning types are closely related, due to how cognitive learning must be based on the behavior of the one who learns (Mayer, 2012, p. 594). Focusing on organizational learning entails seeing learning as an ideal act in a larger structure, and not only something that happens amongst individual staff members. The organizational perspective in this thesis therefore qualifies a holistic perspective, where learning is seen as systematically overarching and not reduced to only the individual level (Risjord, 2014, pp. 122-123). This is based on the sociology of Émile Durkheim and the social facts he proposes, which classify conduct or thought that hold controlling and coercive power outside individual consciousness (Durkheim & Catlin, 1964, pp. 1-4).

Theorists define organizational learning by emphasizing cognition and behavior, where learning encompasses change in both spheres. The "Handbook of Organizational Learning and Knowledge Management" describes organizational learning as processes of transformation in individual and shared perceptions and action. Organizational learning is both affected by and faced with the institutional aspects of the organization. When learning happens for both individuals and groups in an institutionalized way, it is called organizational learning. The acquired knowledge can then be embedded into routines, systems, cultures, and strategies (Dusya et al., 2011, p. 154).

Argyris and Schön share a distinguished theory on organizational learning, where learning is seen through the social dynamics and possible mismatches between expectations and results. This theory for learning is especially relevant for the development field, due to the argumentation for contentment and different experiences for learning (Johnson & Wilson, 2009, p. 39). This difference in experience is related to the possible mismatch between expectations and goals for development and the actual outcome. Specific goals may not be met, which creates a need to understand why. Exceeding much influence on organizational learning theory, the work of Argyris and Schön is found in several fields including psychology, organizational development, management science, and organizational politics (Lipshitz, 2000). Previous scholars have applied theoretical perspectives from Argyris and Schön when writing about Norwegian development cooperation. It has then been suggested that double-looped learning may be essential in facilitating learning from evaluations, and that this requires them to evaluate at later times to establish usefulness, and to use such experiences at an early stage later on (Samset, 2015). The model on organizational learning will be theoretically applied in this thesis to explain different forms of learning leading to organizational learning in relation to evaluation methods.

2.5 Organizational Learning: Single-Loop and Double-Loop

Argyris and Schön first identified learning loops that define different forms of organizational learning. Single-looped learning refers to a single feedback loop connecting outcomes of actions to organizational strategies. For organizational learning to happen in this way, the discoveries of the learning agents must be embedded in organizational memory and result in defining the maps that encode the organization. Some forms of error correction will demand modification of organizational norms. This is when double-looped learning will be necessary, defining situations where errors cannot be corrected through improving the action that the organization already knows how to do. In a double feedback loop, error detection is connected to strategies for effective performance and to norms that define effective performance (Argyris & Schön, 1978, pp. 18-29).
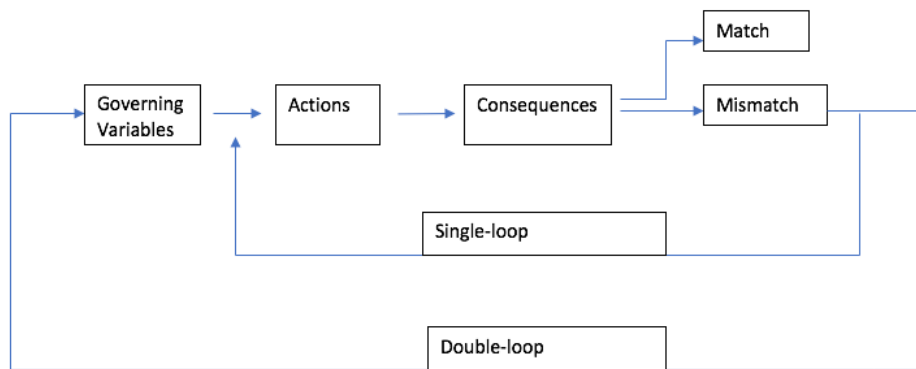
Figure 1: Single-loop and double-loop learning, Argyris, 1999, p. 68

According to this model by Argyris, learning occurs under two conditions: 1. When an organization achieves what was intended, creating a match between the design for action and the outcome. 2. When there is a mismatch between what was designed and the outcomes, and this is corrected. Single-looped learning happens when matches are made, or mismatches are corrected through a change in action. Double-looped learning happens when mismatches are corrected by investigating and changing governing variables. All organizations require single and double-looped learning, where single-looped learning is appropriate for routine, repetitive issues. Double-looped learning is needed for complex problems to ensure the organization's survival. The governing variables of the model are the preferred states individuals strive to satisfy when they are acting, which includes underlying assumptions and values. Learning occurs when actors discover a new problem or a solution to a problem, and when this invented solution is produced. Organizations do not learn, individuals acting as agents for an organization produce behavior that leads to learning. However, organizations have the opportunity to create conditions that influence individual learning (Argyris, 1999, pp. 67-69).

CHAPTER THREE

METHODS

The qualitative research methods used in this thesis will investigate professional opinions on the focus of learning from evaluations in an organizational perspective. Qualitative methods are concerned with intensive studies of small groups of individuals, claiming depth rather than breadth, which gives findings a contextual distinctiveness of the studied experiences (Bryman, 2016, p. 384). The investigation of personal reflections of learning from evaluations is choosen to provide a unique in-depth perspective of the staff in Norad, through theoretical inferences and contextual insight (Bryman, 2016, pp. 399-401). Bryman suggests several quality criteria for qualitative research which includes trustworthiness and transparency (Bryman, 2016, pp. 384-387). This thesis seeks to abide by such criteria though opening up the process under which results are found and conclusions drawn. This chapter will go through the methods applied and decisions made to provide rich answers to the research question and objectives.

3.1 Sampling Method

Specifying the group of participants that will be studied depends on the type of knowledge sought. When studying people in specific settings, such as organizations with their own structures and hierarchies, it is helpful to include people with diverse experiences to include a broader range of backgrounds (Magnusson & Marecek, 2015, pp. 34-36). I was first interested in studying diverse staff as part of the whole aid administration, including both Norad and the MFA, in addition to relevant researchers. I wanted the main part of the participants to be related to Norad through different positions. This is because they hold decision-making power regarding grants and the responsibility for building and disseminating knowledge on development. Therefore, organizational learning from evaluations mostly happens within Norad. I included researchers in the sampling process because they may hold opinions of how evaluations are used in Norad through collaborating with the organization. It would be useful to include participants from the MFA because they represent leadership and make executive decisions on development cooperation. However, I was not successful in

recruiting participants from the MFA, which guided the research towards an organizational perspective only on Norad.

I applied both snowballing and other specific purposive sampling techniques. Purposive sampling is defined as a non-random process aiming to strategically find informants relevant to the research question (Bryman, 2016, p. 408). The inclusion criteria for this strategy were professionals in the field with experience from evaluation processes in different departments in Norad, or through collaborative experiences. I reached out directly by email to several people through this strategy, which was relatively unproblematic, due to all information being public. The snowball strategy was applied by asking two researchers in the field for names to contact, and later asking included participants for people they believed to be relevant (Emerson, 2015). I emphasized asking for people who had different experiences on the topic than themselves. This is important because snowball sampling risks reflecting the researcher's biases through selection within a like-minded network (Loseke, 2017, p. 124). On the first hand, I found during this process that the participants often repeated many of the same names as recommendations, making it difficult to reach out to a large number of people using this strategy. On the other hand, this was also a reassurement that I invited people with a lot of experience with evaluations and learning.

In retrospective reflection, the overall sampling process should have started earlier than it did, seeing as this proved time-consuming. I initiated contact with two researchers in September 2022 to first ask for recommendations on participants. However, the recruitment did not start until December. It later became increasingly difficult to find and contact new participants.

19 people were in total invited to participate in interviews, and eight people accepted the invitations. A requirement for sample size in qualitative research is that the number of participants can support convincing conclusions (Bryman, 2016, pp. 416-417). The goal of such research is to create high quality theoretical inferences based on collected data (Bryman, 2016, p. 399). Eight interviews provided a rich source of material for analysis, due to the diversity of participants. The participants included in the thesis represent a wide range of

backgrounds and experiences that will provide insight from different perspectives. Based on the limited time period of one semester for this thesis, I believe I was able to gather generous information and a comprehensive coverage for the topic that will answer the research question and objectives.

One of the participants is a researcher on international aid; two are currently on staff in leadership positions in Norad, two are currently employed in senior positions in Norad, and three were previously employed in leadership positions in Norad. The chapter for results will not present the specific posistion of each participant that are expressing themselves to secure their anonymity. Meaning that the thesis will not analyze the importance of their specific professional role and experience in the research question and objectives. The results will also not present the participants gender to secure that the answers given will not be possible to connect to a certain person.

3.2 Interview Design

In simple terms, an interview is seen as a conversation with a specific purpose. However, how such a conversation is designed and conducted is not as simple. Qualitative interviews are described by many as a unique social interaction, where this is either seen as an art or a technical skill. The interview is a performance where the interviewer enters as a person seeking specific information, performing through the presentation of oneself and the subject. The participant is cast as someone whose experience matters for the research (Berg & Lune, 2017, pp. 65-66). The researcher must therefore create a situation where the participant can talk about their personal experiences in a way that makes sense for them (Magnusson & Marecek, 2015, p. 46). I personally attempted to create awareness around this to accommodate the participants. As well as familiarizing myself with the participants professional experience before the interviews to gain knowledge of their experiences and worldview, as suggested by Bryman (2016, p. 471).

The interview guide (appendix 5) is based on a semi-structured approach. The guide was created through a dynamic process with several adaptations. The questions of the guide is based on the literature review and theoretical perspectives which serves as a backdrop of

essential topics for the research. The guide includes a list of questions and topics to be covered, while also incorporating a significant degree of freedom in the response and follow-up (probing). In semi-structured interviews, questions may be adapted from the schedule, or new ones may be added based on how the participants respond (Bryman, 2016, p. 468). This method allows interviewers to probe far off-topic and follow the participants toward what they may find necessary to talk about (Lune & Berg, 2017, p. 69). In contrast to standardized interviews, where no adjusting or clarifications of questions is included, I wanted to have the possibility of changing questions along the way because I believe that there may be much information in areas not covered by an interview guide (Lune & Berg, 2017, p. 68). I choose this method for interviewing due to the flexibility it provides when new interesting topics arise that I want to investigate further. While also highlighting the participants own understanding of the topic in depth. Additionally, the research topic had already been developed quite far after the literature review, making the focus of the interview clear and not too general. Starting the investigation with a reasonably clear focus of the research is pointed out as a beneficial situation for conducting semi-structured interviews (Bryman, 2016, p. 469).

The semi-structured interview questions should be open-ended, inviting stories and experiences and reflections of personal opinions (Magnusson & Marecek, 2015, p. 47). The phrasing and order of the questions were flexible an open for interpretation. Following what the participant talked about, we frequently jumped between topics. I used the interview guide to probe about their reflections without steering the conversation. Dividing the questions into sections with an introduction and finish where the most important questions were underlined ensured I had time to cover essential topics. It is recommended to structure the interview guide based on initial open-ended questions, followed by intermediate questions and conclusive questions. Researchers may be interpreted as intrusive by the participants; therefore, a certain rhythm to the conversation may create more comfort with the participant, starting and ending at a normal level (Charmaz, 2001, pp. 679-680). I partly followed such a structure in the guide, beginning the interview with a general question that could provide new insight from the participants point of view. The main section of questions regard their experience with evaluations connected to other thematic areas. I decided to not focus on a certain evaluation type in the interviews to base the thesis on what the participants find

important. This means that the findings and later conclusion is not aimed specifically toward distinct evaluation types, but more toward general perceptions of evaluations.

3.3 Setting and Conducting the Interviews

Eight semi-structured interviews were conducted in January, February and March 2023, constituting this study's main data source. Two interviews were conducted in a private room on the OsloMet campus, four were conducted over online conferences, and two were conducted at workplace locations. The participants chose all locations, and they were all given the option to meet on campus if that was what they found most suitable. One online interview encountered some technical problems, resulting in less time to discuss the subject matter. There may have been drawbacks to conducting online meetings, where we may have lost some face-to-face communication value due to how we were not in the same physical space. Notably, helpful signals through body language between researcher and participant may be lost when conducting online interviews (Brinkmann, 2020, pp. 442-443). Nonetheless, I believe the online meetings provided a rich source of material and was an excellent solution for not being able to meet physically.

In general, the interviews lasted about one hour, and I spent some time after each session to write notes about the setting and other observations or thoughts I had during the interview. Following ethical principles for interviews, I tried to remain as neutral as possible by not leading the participant in a specific direction or being personal about my views (Bryman, 2016, p. 476). I also tried to be present as an active listener, including providing prompts when they seemed hesitant and asking for and giving clarifications when needed (Magnusson & Marecek, 2015, p. 63).

The interview will always be affected by the degree of experience of the interviewer, and I can relate to the stress associated with the first official interviews. Seeing the interview position as a craft, time and adaptation is required to become a good interviewer (Magnusson & Marecek, 2015, p. 67). I conducted trial interviews on two occasions, with a peer and an acquaintance. The goal of the trial interview was to test the timing of the interview, how the questions were understood, and testing the recording equipment. Based on discussions after

the trial interview, many questions or phrasings were changed to better capture what I was attempting to ask. I found that many of the questions I had prepared initially were phrased in a convoluted manner. I further adapted the interview guide after conducting the first two official interviews. This is because the answers I received were not as personal as I preferred, as the participants were hesitant to be specific and spoke in a general matter.

3.4 Ethical Considerations

Some of the primary ethical considerations for qualitative research highlighted in the literature include the minimization of harm, respect for personal autonomy, and preserving privacy. Protection of harm relates to how researchers should avoid damage to their participants as a consequence of the research project. Respect for autonomy relates to protecting the participant's rights and their informed consent to participate in the project. Preservation of privacy involves the degree of personal or sensitive information sought in the research. The principle of confidentiality is also an essential part of securing participants' privacy, where anonymization is perhaps the most known way of doing this (Traianou, 2020, pp. 85-92).

Such ethical considerations have been preserved through avoiding controversial questions in the interviews to the extent that could harm someone's reputation or status. The information sought would also not cause injury to personal or professional environment, due to not covering sensitive topics. All participants received an information letter (appendix 6) to read before the interview concerning their rights and data protection. They have been given a chance to ask questions about this and I informed them briefly on the main points of this letter before the interviews. Based on this, they have all given signed consent to participate. The signed consent forms have been stored in a safe location. All personal information have been handled confidentially, and the names of the participants have been replaced with a number in this thesis as well as excluding any information that could identify the participants. I recorded the interviews using the "Diktafon" app and transcribed them personally as closely as possible to vocal phrasing. This transcription happened directly from nettskjema, meaning that the sound files have not been downloaded or stored on any computer. Transcription documents were made anonymous to ensure that identifiable information is not saved or

included in the research. This project has gone through approval with NSD- Norsk senter for forskningsdata (appendix 3), where both the interview guide and strategy for sampling have been approved.

The relationship between researcher and participant and the power dynamics between them may impact what is said in the situation and the later conclusions. This dynamic may be affected by the researcher's control over what happens during the interview, defining the space where the participant can unfold (Briggs, 2003, p. 500). Therefore, it is significant that the researcher address power relations and engage a process that gives control to the participant to add new perspectives (Gubrium & Koro-Ljungberg, 2005). The role of the researcher and their personal beliefs may also impact the quality of the scientific output. This is a central topic when conducting interviews because the researcher is acting as a tool to collect knowledge. Therefore, it is essential to be aware of ethical guidelines and the value questions that may be part of the research (Kvale et al., 2015, p. 108). I set the agenda for the topic of the interviews. Besides this, I would not say that the interviews constituted any power imbalances. I engaged in processes to give the participants as much control as possible by letting them talk off topic using their own words and language, deciding where to meet, and being respectful of their rights. While also remaining as neutral as possible in the role of the researcher. Based on my impressions, it seemed as though the participants were well aware of research guidelines and ethical standards, as well as being familiar with interview settings.

3.5 Strategy for Analysis

The data for this study has been analyzed through thematic coding. Thematic coding is classified as the most student-friendly approach to qualitative analysis and was chosen based on its fundamental method of answering research questions (Johannessen et al., 2018, pp. 278-280). Nonetheless, the coding happened through a challenging process where several codes overlapped and related to each other across the research objectives. The analysis became an iterative process where the objectives and research question were adapted based on findings in order to base the research on the relevance of themes prominent in the interviews. Thematic coding is based on the outline of Braun & Clarke, who claimed this as a method for analysis, defining it as a rich way to organize and describe data (2006). I decided

to follow the outline of Johannessen et al. due to the openness of the requirements for the thematic analysis and their adaptation to cater to students. This happens in four stages:

1. Preparation: gathering and creating an overview of the transcribed data. This is achieved through listening, reading, and writing notes.

2. Coding: highlighting and categorizing essential points made in the data through close readings of the material and developing this towards further associations.

3. Categorization: sorting the data into greater categories or themes, where each dataset shares characteristics that constitute a common theme.

4. Reporting: to write out the themes and their content in the results to answer the research question (Johannessen et al., 2018, pp. 283-303).

CHAPTER FOUR

RESULTS AND ANALYSIS

This chapter presents empirical findings from the interviews through four themes to answer the research question and objectives. The themes are as follows: 1. Attitudes Toward Evaluation Methods 2. Organizational Factors 3. Political Influences and 4. Evaluation Utilization. The themes provide answers through different angles that were prominent in the interviews and express the connection between evaluative considerations and learning. The review of themes is related to the research question and objectives across each theme.

4.1 Attitudes Towards Evaluation Methods

4.1.1 Relationship Between In-house and Outside Staff

The focus on learning from evaluations connects to how in-house professionals in Norad relate to the process and methods of evaluations. Evaluation research is conducted by outside consultants which makes the relationship with the consultants, and how in-house staff view outside staff important for learning. The participants explained that many consultants are hired through international tenders for centralized evaluations. Which implies that the consultants may not have much prior experience or understanding of the Norwegian development system. Participant 2 and 5 in particular explained the importance of creating trust and spending time with outside consultants to make sure they understand the

Norwegian system during centralized evaluations. In such evaluations, the in-house staff will play a greater role in developing and planning the evaluation as well as educating the consultants. Yet, if the in-house staff experience the consultants negatively, their curiosity and openness towards the later presented findings will decrease. Hiring outside staff to conduct research may be important to secure objectivity. However, this may also lead to insignificant learning opportunities. Participant 5:

> If you experience meeting consultants with a lot of experience and insight into what you are doing, and not just asking idiotic questions, you will be more curious to learn from it and be open for their advice and recommendations […] Clearly, the consultants are very variable. And some are really good in driving processes, and people experience that they are included and heard and understood especially. But others, either if it is personality or lacking insight into something they will be rapidly abandoned, they are not interesting partners for discussion.

Participant 7 described the relationship between the consultants and the in-house staff as a close and trustworthy process. However, others pointed out how outside consultants sometimes are experienced as non-trustworthy, due to how they conduct research. Participant 5 exemplified that the consultants would sometimes be experienced as unprofessional or that they would not be explicit about how interviews would be used toward their participants. This is a reoccurring problem because several participants point to the quality of external evaluators in relation to learning. Participant 1 expressed wishes in relation to planning and conducting decentralized evaluations where it should be made sure that evaluations are conducted in such ways that it is possible to learn from them. And that they should be conducted by competent consultants that can produce something that can be used later on.

It was also pointed out by another participant that if the collaboration between in-house and outside staff is not a trustworthy and confident collaboration, in-house staff will often question the validity of the findings and be less open to learning. Especially if the evaluation findings are experienced as too critical, the in-house staff may see this as unreasonable and be more prone to question its methodological foundations. Participant 2: "If you get a lot of critique you can say `okey this is learning´, at the same time, you become quite demotivated by it, and you can find many arguments as to why this is bad work."

An important experience within in-house staff is that outside consultants may create recommendations based on findings that are difficult to implement in the Norwegian development system. Participant 5: "The consultants will see it from that in an ideal world it should be like this, but this is not the real politics of it." This may be explained through how outside consultants do not have a well enough understanding of the system to provide recommendations for, or that the evaluation is not shaped towards the correct formal goals. Over time, this may have led to a lack of confidence in the outside consultants. Where the recommendations they provide happen in ideal settings, which may result in less openness to learning. This also makes it easier to point blame towards the methods used when findings are experienced as critique. Making it more difficult for the in-house staff to learn and absorb the findings of the evaluation into the organizational system, because they may be closed off to the report based on prior experiences with external staff.

4.1.2 The Probability of Findings

Some methods applied in evaluation research are believed to limit their opportunities to be used widely for learning. This was presented as related to decentralized evaluations through showing dissatisfaction regarding their quality. Many of these types of evaluations are centered around what is called "key informant interviews", conducted by outside consultants, which cannot give answers to broad questions, participant 1:

> When you have a consultant team whose method is to run around and ask key informants about how they think this works, now I am being a little mean and painting with a very broad stroke, you will rarely get concrete information about what works and not.

Participant 1 explained that decentralized evaluations will be typically used by case-workers in the same area when entering new deals and partnerships with organizations to see how their work has been in the past. Furthermore, that it is the bad quality of such evaluations that makes them less operative across the administrative system. The lack of utilization across the organization may be inhibiting for double-loop learning opportunities. Due to how evaluations are not believed to be good enough to be shared with other areas or thematic sections, they will not take part in building knowledge that can create large changes. Findings

from one decentralized evaluation may be relevant across the organization and take part in general increases of knowledge. However, they are not being utilized as such, which limits opportunities to only provide single-loop learning. However, participant 1 also explained that the bad rated quality of decentralized evaluations, together with a general interest from management to increase learning, may have played a role in forming the current signals to prioritize impact evaluations of development cooperation programs in the aid budget.

Participants 1, 2 and 8 expressed that because many parts of development cooperation is difficult to measure and evaluate, much of the knowledge is based on probable findings that cannot provide sure conclusions. Participant 8:

> Some assumptions are almost facts, like vaccines, for example the polio vaccine, if you get the needle in your arm then you won't get polio that's pretty safe […] womens rights can be very complicated, then it will be like we know that in some societies this have helped, and we will do it again in another society because we think it will help. You will be able to say something about the development, but we cannot say it precisely like you can with vaccines which you can count, or study admission spots at universities, or children finishing basic education.

Findings based on assumptions may be challenging to generalize into something larger. Participant 1 call themselves hopeful that if the general quality of evaluations is increased, this will in time contribute to a larger understanding of the results of development cooperation and make the evaluations more operative. Some believe this is best answered through impact evaluations. However, impact evaluations is not currently used widely in Norad due to the costs and ethical considerations one has to take with regards to control groups. As participant 5 pointed out, using randomized control groups will mean that development aid is restricted from some people, which is a difficult decision to make. Participant 6 mentioned how impact evaluations are not seen as something that can and should be conducted often in Norad. Whereas participant 5 explained how they did not believe that impact evaluations were suitable for all areas of development, and participant 7 explained how established impact may be much more relevant in some contexts than others which is a limitation. This points to inconsistencies in how impact evaluations are considered and may explain why it is not used widely.

4.1.3 Dissemination of Findings to Promote Learning

Some staff members in Norad take on a great deal of extra work when disseminating knowledge from evaluation findings. This is done through many different outlets. The dissemination of knowledge is partly dependent on the personal beliefs within staff, where some have worked more towards the specific goal of learning than others based on how important they consider this to be. This means that to some degree, learning opportunities will vary together with whoever is in charge of spreading evaluation findings at the time. According to participant 3, efforts are needed to make sure that policymakers are aware of evaluation findings and to create public debates. Participant 2 expressed strategies for dissemination to facilitate more inclusive processes:

> It has happened multiple times when those in charge of a program were dissatisfied after the first draft came, but that was maybe not strange because those that had worked on it were not very familiar with Norwegian aid. Then we had a process where you had quite good discussions between those that evaluated and those that worked with it and at last the section was happy with the results. This was not because it became less critical, however, because they felt that they were a part of it, had some ownership over it. That's why I think because of learning situations there should be more of these contributory evaluations.

Other strategies pointed out that were applied to mediate and push findings towards policymaker's attention include creating an outside debate to include the political sphere. Participant 3 elaborated deeply on all the extra work done to disseminate findings. Using international networks for reference groups that would comment and discuss evaluations, hosting seminars to launch reports where relevant politicians, researchers and journalists that would write stories of the findings were invited. The point of such strategies was to create a knowledge-based debate around the centralized evaluations that would gradually improve development cooperation. Others also mentioned the importance of public debate around evaluation findings as this may be seen as a form or signal of control explained by participant 5. Participant 3 explained in a humoristic tone how difficult it was to get certain people on board the public debate, where they encountered many hindrances: "The Ministry of Foreign Affairs were representing political management, so they often tried to stop these things, they did not want debate."

Participant 7 also mentioned the importance of hosting seminars to present centralized evaluation findings. However, not emphasizing dissemination strategies, debates and learning as deeply. This shows ambiguity in dissemination strategies where learning opportunities may be dependent on whomever takes on the task of disseminating findings. Familiarity with research findings regarding grander thematic areas as presented by the centralized evaluations may be important to facilitate learning. If one is not aware of what research has been done or what it concludes, one will not use it in their work.

4.2 Organizational Factors

4.2.1 Rhetoric Focus on Learning

Some believe that the focus on facts and learning in Norad merely composes a rhetoric function. If Norad was a learning organization, participant 6 believes that the recourses would have been allocated differently to actually support and prioritize facts and learning. Participant 1 explained that signals to prioritize knowledge coming from top management may become important for case workers and other staff over time. However, participant 5 explained that in a busy everyday work life they currently have no choice but to prioritize differently. As expressed, evaluations are not something you really think about (if you work outside the evaluation department), unless it is your work that is being evaluated, participant 5:

> [...] but most here, in reality, will not relate to evaluations on the daily anyways. It is more if you are involved in an evaluation or your area is being evaluated and if you are initiating something new where you think you need to see what has been done previously and what have we learned from it

Participant 8 explains that all agree learning is important and share an interest in doing the best work possible, except there is no consensus on what this means in practice and people may have a distanced relationship to evaluations: "When it comes to evaluations in general, many have a distanced relationship to them and they are a little scared that they will come to our area [...] It is something everyone agrees is sensible, just not in my area." This shows disparities between opinions on learning, where participants explain that the leadership emphasizes knowledge, without creating a space where staff can relate to this. This is

apparent in the increased workload for all employees in recent years, making it very difficult to account for learning or using evaluations in their positions, as participant 4 explains:

> There has not been an increase in positions, the budget is growing, the responsibility is growing, but they are close to a freeze in positions. It is on the same level today as it was 20 years ago, while the aid budget, and the responsibility of Norad has increased four to five times. This gives consequences on the ability to follow up.

Evaluations are sometimes not considered doing because it is seen as a misuse of money that could have been spent on projects, speaking to a lack of focus on learning and being more focused on doing. This is understandable based on the ideological background of development cooperation, as explained by participant 8. The staff wishes that the money they provide would increase the welfare of others. However, they are not as focused on the results of this. Participant 8: "It is often that the driving force to help is stronger than the driving force to see if it did help, did it get better? I often think this is lacking, at least there is not enough of it."

The focus on knowledge and facts is not a new phenomenon, participant 4 explain that this has been talked about in general public administration and in Norad for many years. Despite calling this a rhetoric function that is not prioritized amongst staff, some measures have been taken that could support double-loop learning. Several participants pointed to the recently started department for knowledge when talking about learning and evaluations. Work done in this department surrounds collecting, distributing and utilizing knowledge, according to participant 8. This re-organizing of the organization is an example of potential for double-looped organizational learning where they attempt to improve the organization through new measures. Perhaps also changing the organizational norms on how knowledge is viewed in the future. This is a rather new department, it is therefore not possible to conclude on its effect yet. However, most participants were very optimistic regarding this department.

4.2.2 Cultural Aspects

There is no apparent culture for facts and learning among the organization as a whole. Participant 6 explained how they experienced that people did not always think twice about if their work methods were the best option in all circumstances: "You should pretend to know,

and that is the same culture as everywhere, so it is like you fake it a little bit. And that is totally opposite towards seeking knowledge." This perpetuates a focus not geared towards openness to learning. Having an open culture for learning may be important, where participants propose that it is vital that this starts with the leaders. Participant 8 exemplifies this:

> If the director says that we need to follow up on this, then it will be followed up. And this may go down gradually through department directors saying that we need to follow up on this and it will be followed up. But if it is only case-workers saying that this is important and good, but only I am interested in it, it will be put away in my drawer.

Some experiences resistance from top management when trying to assert more openness in the organization through evaluations. Participant 6 described a feeling of not having enough time to fight for changes in Norad seen as necessary by the participant, after several people attempted to stop an evaluation they proposed. This builds under the rhetoric focus on evaluations, where in order to gain larger learning opportunities, the people in management and political positions have to make this a priority in the organizational culture.

4.3 Political Influences

4.3.1 Perceived Boundaries

Learning happens once we become aware about what works or not. However, development projects take time. The official development aid from Norway is given within a highly politized field. There are several things politically decided that may be experienced as having impact on possible learning opportunities, this includes boundaries in time and recourses. Participant 1 explained that the job done in the thematic sections in Norad is acting as operators of political goals, where they take the broad goals stated by politicians and put them into practice. The professionals follow what there is political interest for. This is highly relevant for evaluations based on the fact that one needs to evaluate what is politically relevant in order for the evaluation to be used later on. Participant 5:

> Sometimes we experience that the evaluations came a little too late, because it takes a long time, maybe there is a new government in the meantime and the priorities have changed. Then it is experienced as less relevant somehow.

Political changes of priorities are important for evaluations because evaluations will not be considered important enough to guide such priorities. In other words, the politics will trump what the research says. Participant 1 exemplifies that if the Norwegian government has decided that we should be an important actor in peace processes in a certain country, this will happen no matter what an evaluation says about the subject. This also applies for swift changes in political interests either by the changing of governments or unexpected situations occurring. Participant 2 talked about the importance of keeping long term perspectives in aid. This may be difficult due to how different governments may prioritize different areas. New situations means that politicians wants things to happen fast without much prior information, participant 5:

> […] political priorities change. And when the politicians first have decided for something it has to go very fast. So, then we think we know what to do usually, but there are reasons that you consider that we don´t have time for this if you want to get something going in six months you cannot do thorough analyses of the countries you´re entering.

participant 8 expressed thoughts on the uncertainty related to new situations:

> For example, we are now waiting to know what happens with the Ukraine funds, which are going to increase drastically, what then happens to other aid? They say they are not going to reduce it, but what about the capacity to give that aid, are they going to increase that? This we do not know, so it may happen that one's time is even more shortened.

This shows a lack of knowledge based thinking from political management, setting boundaries for learning within Norad. Additionally, case workers have to abide by rather strict economic rules that may make adaptations based on new knowledge difficult. This does not create for a learning system because abiding by such restriction means that they are unable to work based on what they learn from the evaluations. Participant 8:

> A project can cost 500 million kroners, and you will have 100 million per year, and you have to pay out those 100 million that year, you could not pay out 300 million and see some years later how that money is doing because you will not have that money then. It will always be 100 million if the government has granted it in such a way. And it is not always that fits as good with the progress of the project which is a factor that can lead to bad time and bad control.

Political constraints in the aid system inhibits both single and double-loop learning through not creating openness or space for changing organizational norms, or the power Norad staff has to make changes for efficiency due to budgetary decisions. Development professionals are interested in helping people, while the larger development administration is also part of the political sphere. This does not seem to provide good foundations for learning due to overbearing political influences.

4.3.2 Evaluative Independence

Although the centralized evaluations conducted by the department for evaluation abide strongly by the principle of objectivity through hiring external consultants, the department are still deciding what to evaluate centrally. This is not necessarily an objective process. Some highlighted the importance of evaluating what is politically relevant at the moment, and how this plays a part in getting the evaluation utilized. This happens within an interesting field where the evaluation department should be independent. However, in practice somewhat following what the political management wants to see, participant 3:

> The largest prerequisite for use is that something is relevant for those that will use it. And that is often about something that is relevant today. If you as an evaluator for example in Norad or the Ministry of Foreign Affairs, could know what is on the agenda one year from now and come in with and evaluation on that, so you need to be good at predicting.

There may be many important areas to evaluate which are not at the political forefront. The department for evaluation have final say when it comes to thematic centralized evaluations. However, they also need to facilitate interest in their work which means abiding to the political climate. Showing that evaluations do not have the weight to change political priorities. To some degree, learning is dependent on what political management wants knowledge of.

Participant 3 explained practices of conducting meetings with the MFA and the MCE, where they could advise what they were interested in gaining new knowledge about. On the first hand, including users of knowledge into the evaluation process may facilitate greater use later

on. On the other hand, it may remove some objectivity from the independent evaluation department. Where the evaluations may be influenced by what others find important, and not necessarily what the independent department finds important.

4.4 Evaluation Utilization

4.4.1 Focus on Accountability

Evaluations do not only have an effect within the organization, some participants believe that the importance of evaluations are their function outwards towards the public eye. They can be used to signal what type of organization Norad is and that they have the ability to be self-critical. The mandate for centralized evaluations is dual, where they are supposed to serve the function of both learning and accountability on the basis of public spending. Assumably, showing accountability or control through evaluations may also lead to learning due to outside debate and interest in how the funds are managed, which in some cases have given pressure to changing ways of working. An example of this is brought forth by participant 3, of a controversial centralized evaluation, where many instances tried to stop this report being published because it showed that the quality of an external organizations work in a certain programe was below par. The report was published despite this and created a lot of debate, where the external organization intensively tried to defend themselves. According to participant 3, after a while this evaluation contributed to the re-thinking of working ways and norms within the external organization when they realized how useful the evaluation was and it also led to the MFA following up on this. Which in the end created learning opportunities for both an external organization and Norad through shining light on new aspects of a project.

Participant 5 argued that focusing on learning as the main priority may be too difficult, this may be based on their expression of how many evaluations show the same recommendations time and time again without this leading to anything. Participant 5:

> I think it is important to communicate outwards to the public that we have a type of role where we look ourselves in our cards, and then it is demanding to achieve learning […] I think there is a very important function clearly outwards, to show that we have an external look at ourselves in addition to Riksrevisjonen and that type of stuff.

Participant 2 explained control through managing public opinion because there are a lot of questions raised about development aid, partly because outcomes cannot be seen or experienced directly because things happen so far away: "There is an absolute function especially because there is so much writing about questions of the aid." This points towards some believing that outside control and public image being just as, or more important than focusing on learning. However, that learning may possibly take part in this process. The outside control through the public may be part of creating importance around evaluation findings, forcing the organization to change when facing complex issues.

4.4.2 Single-loped Learning Presence

The participants used many examples that would classify as single-looped learning from evaluations, which were often dependent on the surrounding methodology for the evaluation. This puts single-loop learning as the most frequently experienced form of learning, and if evaluations seem irrelevant they will often lead to nothing and be disregarded. This is especially prevalent for decentralized evaluations that do not have the weight to be used several times for multiple purposes. Participant 4:

> The most important for learning from evaluations is that they contribute to, or should contribute to, that we know what works and therefore can adjust the course and adjust projects so that we get most worth from the money. And that they can contribute to reaching the goals of the project.

This form of single-looped learning from evaluations can take the form of solutions directly feeding into new actions, if the evaluation is timed correctly. It can also take the form of changing actions based on mismatches intended to better achieve goals. Participant 2:

> One example I remember best from the educational sector in an Asian country, there were many givers, and we had a forum for givers that went through different sessions. One year, I think Norad had the main responsibility there when we were supposed to evaluate the last phase of the program to plan the next and we were asked to contribute to this. We did this through an international team, which we managed to do rather quick and thoroughly. The evaluation then went directly into the planning of the next phase. And that is learning or follow-up which is very good, usually the evaluation would not come at that time.

Participant 8 explained practices within micro financing:

> And the point of that was to give money to small businesses, they could loan money to run the business and pay back the money, this was a large success. But the success was measured in that the money was paid back, and that they at least did not disappear, which is a good indicator. In later years it is believed that giving money instead of loaning money works better in some areas. Not all areas, but those connected to questions of individual and household level poverty. So, for a family that does not have money, giving money is some of the most effective you can do to secure education, health and hunger. And that is quite a large change, which is based on research and controlled studies […] and based on cash transfers we have seen that this often has an effect on many other areas. Where you then should see, does this have larger effect than the solution we have choosen, should you if not choose a cash transfer. So, this is often used as a benchmark in the industry in some areas.

Others experience this form of single-looped learning as restrictive, where it is difficult to use knowledge when starting something new when previous evaluations have only been created with specific adjustments in mind. Most evaluations, besides the yearly thematic reports from the department for evaluation will not be made to generalize to other areas based on findings. Participant 7 and 8 explains that this means that for example working with cash transfers in an area with high corruption or war will not be as effective as in other areas. Or that certain norms or local regulations in other areas will not make it possible to achieve the same goals. Meaning that programmes and case-workers will need great awareness of the locality of where the aid is given and may create a negative association surrounding evaluations because most will not be relevant for them. Other times, programmes that have been effective one place will be tested in new places in the hopes of having effects in new societies. In other words, there is a lot of uncertainty in development cooperation and it is rather difficult to learn something as a hard fact.

Some believe that the professionals within development aid wants to utilize knowledge, but that they experience that this is not readily available. Participant 8 explained that systematically, there is no way to ensure knowledge storing and using, and people do not know where to turn for the latest evaluation. This is a large part of what is currently being worked on in the department for knowledge, through creating systems and technological services that would support knowledge utilization.

There are some systematic efforts seen that possibly would ensure learning between Norad and the responsible state department as explained by several participants. Where the recommendations from the centralized evaluations would be sent to either the MFA or the MCE and then the responsible department would within eight weeks respond to the foreign affairs secretary general on how they were to follow up on these recommendations. After that they were to report on what had been done after one year. This is a systematic means of action to attempt to ensure change and learning. However, there is no systematic security that this will lead to anything. The political management will perhaps have different priorities than the evaluators or Norad, meaning that this may not ensure action change.

CHAPTER FIVE

DISCUSSION

Going in-depth on learning from evaluations within Norwegian development cooperation shows how complex and intertwined this topic is. This topic was larger and more multifaceted than anticipated, which proved challenging in analytical presentation. This also reflects the experienced complications of learning from evaluations. However, these nuances provide for interesting discussions in relation to the theoretical perspectives. The synthetic theoretical perspective will guide an understanding of the research question and show the way for future research on such topics in similar organizational contexts. The discussion applies theoretical concepts presented in chapter two from Weiss (1998a), Argyris & Schön (1978), and Argyris (1999).

This chapter will first discuss essential answers to the research question and objectives across the four presented themes. The objectives overlap and relate across topics presented by different themes in the discussions of empirical findings. The most comprehensive perspectives are related to research objectives 1, 2, and 4, because this was most prevalent in the data which means that such considerations were most important for the participants. Objective 3 finds less relevance in the discussion because the collected data is not able to provide a comprehensive answer to this. The discussion also presents a debate relating

empirical findings to previous research on the topic. Lastly, it presents a discussion of the relevance of the theoretical perspective.

5.1 Attitudes Toward Evaluation Methods

Attitudes toward evaluation methods provides perspectives that overlap objective 1 and 2. Objective 1 consider attitudes toward usefulness of evaluations, while objective 2 consider how methodological considerations in evaluations can affect learning possibilities. How professionals consider usefulness of evaluations are partly seen through the quality of work conducted by external consultants. Evaluations will be considered as less useful if this quality is lacking, resulting in less openness to learning. Such attitudes are therefore partially reliant on chosen methods and the balance of in-house versus outside staff involvement.

The relationship between external consultants as outside staff and the in-house staff in Norad is central in the explanation as to why staff perceives the evaluations to be of too bad quality to provide learning. This provides a dual perspective on learning. First, the outside consultants are perceived to not be qualified enough to provide sound research which can be incorporated in the Norwegian development cooperation system, thus resulting in negative associations towards learning from evaluations. The blame of the unreliable methods in evaluations is often pointed toward the external consultants. When at the same time, it is the responsibility of the department for evaluation to hire the consultants in centralized evaluations. Second, due to experiences over time, evaluations that are experienced as negative critique toward staff may be disregarded on a methodological basis. This means that individual and organizational learning may be avoided, due to the lack of administrative confidence in outside staff on the basis of determining evaluation results as critique. With no present system or ideas to improve the quality of evaluations through the confidence in outside staff, the methodological basis and relationship with consultants lead to less opportunities to learn.

Considering different evaluation methods that affect use and learning opportunities is also seen through centralized and decentralized evaluations. Decentralized evaluations show most potential for single-looped learning connecting mismatches and actions taken by those

ordering the evaluation. Such evaluations may only provide the basis to account for effective solutions through single-looped learning, based on specific information which may embed the organization (Argyris, 1999, p. 69). Decentralized evaluations were pointed out to only have potential to be used by those ordering it, and not to provide any lessons that may travel through the system. Such unconvincing evaluations therefore inhibit double-looped learning through not taking part in processes that would increase overall knowledge and perhaps call for greater change. The centralized evaluations show potential for double-loop learning, due to dissemination strategies that may pressure the organization to learn and move forward. The empirical findings pointed to dissemination strategies serving a form of accountability through informing the public and the political field of evaluation findings. The inclusion of potential knowledge users in the dissemination process may be important to facilitate learning in the organization, and this can happen with in-house staff or political management.

5.2 Organizational Factors

Identified organizational factors in Norad provide perspectives on objective 4, which seek an understanding of how such factors may affect attitudes toward evaluations. Attitudes toward evaluations are influenced by the fact that there is no real system or structure that prioritizes learning. This creates a lack of a shared methods to ensure that learning happens, and that evaluations are part of this learning. The results for this thesis emphasize a focus on doing in Norad and an eagerness to manage development programs and budgets efficiently. However, this focus may take away time and mental space from learning. This is partly explained through the large workload on staff and the restricted time this leads to. The focus on knowledge from management is believed by some to constitute a rhetorical focus due to how it is not made a priority amongst staff through facilitating arenas for learning from evaluations. All may agree that both learning, and evaluations are important parts of development cooperation. However, there is a coexisting uncertainty concerning how this learning should happen.

According to Weiss, creating new strategies for learning or adaptation is difficult, due to the recourses needed to move away from the current status quo. In order to create circumstances that would support transitions one needs highly convincing research suggesting

that new strategies would be beneficial for the organization (Weiss, 1998a, pp. 310-314). Evaluations are often not seen by the participants for this thesis as convincing enough to provide a basis for learning or organizational transitions, due to beliefs of lacking methodology and administrative confidence. This means that the general perception of staff in Norad point to disbelief in evaluations as convincing enough to provide double-looped learning.

Several perspectives provide a view where learning is unattainable in practice, which leads to habits of continuing doing what they have done in the past and not being open to question this. Such factors in the status quo would call for increased efforts in dissemination practices and evaluation cogency by Weiss (1998a, pp. 314-317). However, dissemination aimed toward learning is difficult because interpretations of this varies within staff which means that opportunities to learn of evaluation findings will also differ. Creating more cogent evaluations will nonetheless be a factor in increasing learning so that professionals may feel more secure in trusting its findings.

5.3 Political Influences

Political influences relates across research objectives 1 and 4. Objective 1 concern attitudes toward the usefulness of evaluations, and objective 4 consider organizational factors that may impact attitudes toward evaluations. Organizational factors may impact attitudes toward evaluation through how the participants explained that political decisions may limit potential for evaluation and learning. This also relates to how the staff often do not have time to learn and how this is not prioritized in practice. Additionally, if an evaluation is not seen as relevant, this will impact attitudes towards how useful it is regarded. Whereas the evaluations that answer to current affairs and needs are those that are centrally regarded as the most important and useful.

Being an organization that practices and provides for organizational learning while also being subject to governmental ministries may provide conflicting interests. This conflict is based on perceptions of the hindering of learning through recourse allocation and budgeting, and the reliance of political relevancy for utilization in centralized evaluations. This thesis also presents perspectives where staff experience resistance from political management in

conducting and disseminating evaluations, pointing to difficulties in navigating the administrative system. To be able to provide cogent evaluations that are also disseminated and utilized for learning, this needs to be prioritized throughout the whole system amongst all staff and such a change has to be specified from top management.

Evaluating with political awareness is recommended through strategies for evaluation outlined by Weiss, which Norad follow closely (Weiss, 1998a, pp. 314-417). Utilization of centralized evaluations is heavily reliant on political relevancy, and evaluations are often shaped based on what there is interest for. This means that Norad prioritizes objectivity in research by hiring external consultant to conduct evaluations. However, decision-making of what to evaluate thematically is not necessarily such an objective process and is not always decided by the independent department for evaluation alone. Learning is therefore connected to current political relevance. This can create difficult situations due to the fast changes in international politics. New situations and needs may present themselves unexpectedly, such as concretely experienced by staff with the breakout of war.

5.4 Evaluation Utilization

Evaluation utilization is connected to learning, due to how facilitating use may also lead to learning. Attitudes toward utilization of evaluations stipulates a perspective on research objective 2 concerning how methodological considerations in evaluations may affect learning. Different methods for evaluation will both affect attitudes towards evaluation utilization and the organizational learning that may happen as a consequence. If the methods for evaluation are seen as poor, which was mainly pointed out in decentralized evaluations, this is seen to result in single-loop learning. Where the weight of evidence is greater as in meta-evaluations, they may potentially provide for double-loop learning. Research objective 3 consider how evaluations may be seen as necessary for changes. This thesis gives perspectives where evaluative research with a great sum of evidence may take part in changes where evaluations are considered as useful. Evaluations may not be seen as necessary for changes. However, they may become important during transitions to see where efforts should be made by the new department for knowledge.

Centrally conducted evaluations share two mandates, where both learning and accountability should be provided through the evaluations. This thesis finds that the mandate for accountability can also be seen through managing public perceptions of Norad as an organization that is able to practice self-criticism and provide information about the outcomes of development cooperation. Learning can take part in this process through dissemination strategies that spread findings or include users of knowledge. However, this does not necessarily happen easily or straightforward due to different interpretations of this among staff. The empirical findings present attitudes where managing learning is too difficult because there is no systematic efforts for this across the organization. This may result in accountability becoming the main priority for centralized evaluations.

Meta-evaluations that overarchingly evaluate the quality of evaluations provide greater weight of evidence and have resulted in prioritizing learning in Norad. The empirical findings show that meta-evaluations that point to a lacking quality in decentralized evaluations may have created a shift in staff perception. This is because participants used the quality of decentralized evaluations to explain the current state of knowledge within the organization and where they need to lay efforts to increase quality and learning. This points to the importance of credible evaluation findings, which may be based on meta perspectives that pool a larger weight of evidence in decision-making processes (Weiss, 1998a, pp. 314-417).

The department for knowledge in Norad will have chances to learn from such meta-evaluations in their work to systematize knowledge and incorporate this in the rest of the organization. Incorporating this department into the work done in other departments in Norad may be highly beneficial for organizational development. This may create double-loop learning if the department for knowledge alter perceptions of staff and strategies used. Which can happen through changing governing variables of values and actions. Once such solutions are produced, they have accomplished double-loop organizational learning (Argyris, 1999, pp. 68-70). Despite this, this thesis finds that single-loop learning is currently most present in Norad. This is partly because of the lack of confidence the staff has in evaluation methodology and findings. Which creates situations where single-looped learning is believed to be the best an evaluation can provide. Especially decentralized evaluations that are used for effective

changes in actions based on evaluation findings, resulting in a lack of circumstances for double-loop learning.

5.5 Findings of the Thesis in Relation to Previous Research

Previous scholars suggest facilitating learning through communication for technical knowledge, and through involvement for wider issues in policy. In such ways it is suggested that the evaluative system may help the organization understand their basic assumptions and see necessary changes (Forss et al., 1994). Organizational factors that concern inclusion or exclusion and communication among the staff is also relevant in this thesis. However, empirical findings of the interviews proved more significant towards professional impressions of externally hired consultants and their invoked administrative confidence. In light of theoretical guidance, this thesis adds the perspective of the importance of the relationship between in-house and outside staff to facilitate sound evaluation methods and potential learning. Additionally, this thesis includes the benefits of facilitating involvement through dissemination strategies where staff can feel on-board with evaluation findings.

Previous work point to both organizational factors and methodological considerations in evaluations to explain utilization and learning. A study shows that evaluations are not amongst the top arenas for learning in Norad, because other domains of work may facilitate this more naturally (Samset, 2015; Samset et al., 1993, p. 64). Other scholars present how the organizational capacity in Norad to receive technical information is much larger than the capacity to learn abstract knowledge (Forss et al., 1994). Looking into methodological decisions and challenges experienced when evaluating programs shows difficulties of linking cause and effect in dynamic human societies. Studies explain that evaluations often encounter challenges in identifying impact, due to seeking impact where this is not possible to provide evidence of, separating between impact of aid and other components, and applying poorly developed methodology (Villanger & Jerve, 2009). Decentralized evaluations in particular prove to be lacking through their methodological bases, and are mostly used instrumentally to inform ongoing projects (Chapman et al., 2017).

Following views presented on organizational factors related to learning and methodological considerations in evaluations, this thesis connects such views in findings stating that evaluative methods are significant in learning processes. Empirical findings of this thesis finds that evaluations are often not utilized for learning, due to the underlying methodological basis of evaluations that staff find to be of bad quality. Because staff care about how evaluations are conducted, this will be considered in decisions of utilization and learning. In support of previous findings, this analysis found that decentralized evaluations served purposes of single-looped learning classifying action changes in programs. However, that such meta-evaluations as conducted by Chapman et al. (2017) provide much larger weigh in adjustment processes.

This thesis finds that learning opportunities may be related to dissemination strategies. This can be seen in light of the debate on prioritizing accountability or learning through evaluations where previous studies have concluded that there may be a trade-off between the two that is necessary to acknowledge (Reinertsen et al., 2022). Others believe that the methods needed for the different goals are too different to combine and that this may create tension (Armytage, 2011). This thesis finds accountability through spreading evaluation findings that promotes external control while also providing opportunities to learn. This form of accountability is not to be confused with methods for conducting evaluations with a focus on controlling public spending. Accountability may be interpreted through dissemination strategies that ensures follow-up on findings. Through cooperating with the public, such dissemination may hold Norad and political management accountable for evaluation findings. Learning may take part on this through increasing awareness, or by pressurizing changes in action.

Previous scholars have suggested that it is important to create double-loop learning through evaluating at later times in programs to establish effect. The findings should be used before planning new programs to correct foundational assumptions (Samset, 2015). Such practices may be essential in creating opportunities to reach goals stated in Norad´s strategy towards 2030, where knowledge utilization and learning has been set on the agenda (Norad, 2021). This thesis adds the perspective on importance of also creating evaluation methods

that lays foundations for learning. The empirical findings show that in order to facilitate single-loop learning, but also double-loop learning, evaluations should be conducted based on high standards and shared systematically across the organization. In addition to this, organizational components should strive to utilize evaluations more systematically.

As this thesis emphasizes, participants believe that the focus within Norad is on doing, and not as much on learning about outcomes. This factor supports previous work looking into the motivation behind the work done in the Norwegian development cooperation system. It has been found previously that avoiding learning about mismatches in programmes creates abilities to constantly implement new projects instead of using time and resources on learning (Lindkvist & Dixon, 2014). This thesis sees such factors in relation to political collaboration and the effects of this in Norad. Restrictions in time and resources allocated towards learning, results in staff feeling as though political management emphasize learning without creating space for it.

5.6 Discussion of Theoretical Relevance

The contribution this thesis makes to existing knowledge is the establishment of a link between evaluation methods and organizational learning, through analyzing professional attitudes on the topic. The empirical findings were successful in establishing a synthesis between considerations of in-house versus outside evaluations and dissemination and utilization theory outlined by Weiss (1998a), with learning loops theorized by Argyris and Schön (1978) and modelled by Argyris (1999).

Objectivity is ideal in evaluation research and the in-house versus outside debate (Weiss, 1998a, pp. 37-38). Norad abides by this through using external consultants in evaluations. The findings show that this practice may impact learning both positively and negatively. Positively in the sense that it is important for staff in Norad to be able to trust in reliable evaluation findings. Negatively in the sense that abiding by such objective standards may result in less learning because in-house staff is not involved in the process and may be unaware of its findings. Relationships that regard administrative confidence in external evaluators were found as highly relevant for learning, where this is seen to impact the openness the staff has

to learn the lessons provided in the evaluation. If the in-house staff consider the external staff as unprofessional or the methods they use as unreliable, they are not interested in the findings, resulting in limitations of both single-loop and double-looped learning.

The relationship between in-house and outside staff also connects to how external staff understand the program under evaluation. Negative experiences of unrealistic recommendations from external evaluators may lead to beliefs that their findings are difficult to incorporate in the Norwegian setting. The considerations for in-house versus outside evaluations may impact both single-loop and double-loop learning. However, the empirical findings of this thesis are more aimed towards the opportunities the considerations provide for learning for in-house staff.

Dissemination of evaluations are important to secure attention of key staff which is facilitated through including knowledge users (Weiss, 1998a, pp. 305-307). The empirical findings show that this connects to organizational learning through in-house staff pushing evaluation findings in internal seminars, through the public sphere, or by inclusion, to change actions or call for larger normative changes in development cooperation. Dissemination has shown to provide opportunities for learning, through creating openness toward evaluation findings and pressuring action changes. However, it is uncertain whether this results in single-loop or double-loop learning. Learning based on dissemination strategies may be restricted by the ideal of objectivity because there is limits to the degrees that in-house staff can be involved in the evaluation, which may make it easier to disregard the report.

The empirical findings present that the prevalent status quo of Norad is proved as difficult to disrupt based on learning from evaluations. The findings supports theory on the fact that large amounts of cogent research is needed to result in learning. Both political constraints and organizational characteristics may define the need for large amounts of research (Weiss, 1998a, pp. 310-317). However, meta-evaluations may take part in large changes happening in the department for knowledge that may classify as double-loop, and single-loop learning. This underlines the important connection between sound evaluation methods and organizational learning.

The most coherent connection between evaluations and learning are seen through decentralized evaluations. The empirical findings clearly shows that such evaluations are used for single-loop learning by staff managing development programs in Norad. Single-loop learning is error correction that embeds organizational memory through changes in action (Argyris, 1999, p. 69). This happens with decentralized evaluations through defining the work done in case management where incentives are to efficiently work towards selected goals. When evaluation methods increase to meta-studies of decentralized evaluations, this has shown to provide greater opportunity for double-loop learning through the department for knowledge. Centralized evaluations are less coherently answered with regard to learning. However, centralized evaluations are often followed up by dissemination strategies which may increase learning opportunities.

Less relevant connections between the theoretical perspectives and the empirical findings include the autonomy of external evaluators and external dissemination of evaluation findings (Weiss, 1998a, pp. 38-39). The autonomy to provide research that may challenge existing frameworks were not discussed because this was not prevalent in the empirical findings. External evaluators bringing evaluation findings to top management were also not found to be relevant. Dissemination happens through in-house staff, possibly missing chances for greater utilization of evaluations through excluding outside staff in this process.

Additionally, this synthetic view does not fully account for the different types of learning loops that happens from evaluations in practice because this was not always possible to extract from the collected data. It was difficult to distinguish levels of organizational learning in practice, due to how the interviews were geared toward chances that facilitate learning more than practical consequences resulting in learning. However, the thesis does account for the opportunities for learning that evaluation lead to, which may result in either single-loop or double-loop learning. A clearer specification of different learning outcomes in loops would require further interviews on the topic to follow up with utilization and learning among staff. Especially in regard to double-loop learning, where the thesis is only able to distinguish opportunities that may facilitate this, but not how it happens in practice. This weakness is also

prevalent in classifying individual learning leading to organizational learning, where the results in this thesis is not comprehensive enough to see all distinctions between the two. However, the perspective of organizational learning is still highly important in understanding learning in development cooperation.

The chosen theoretical perspectives gave insight into professional attitudes on learning from evaluations, facilitated in an organizational perspective. However, the synthetic theoretical perspective is incomplete, due to the perspectives that has been left out. In the end, the discussion would have benefitted from a more comprehensive view, especially by including Caplan´s presentation of the distinctions between researchers and policymakers. The two communities theory explain the non-utilization of research by investigating relationships between researcher and policymakers, who arguably live in separate world with conflicting values (Caplan, 1979). This would provide interesting perspectives in Norwegian development cooperation through an analysis of evaluators and decisionmakers to understand their utilization of evaluations. This theory, and other perspectives presented in chapter two would be interesting to utilize in a re-analysis on this topic. However, Caplan´s theory was not possible to include, due to not having enough participants with backgrounds in research.

CHAPTER SIX

CONCLUSION

This conclusion will briefly summarize the most important perspectives based on the empirical findings to answer the research question and objectives. Before including a discussion of recommendations for future research based on what this thesis leaves out and its restrictions. Lastly, it presents the limitations experienced in the thesis to examine weaknesses of the research.

6.1 Conclusion

The discussion created theoretical inferences between the empirical findings and the theoretical perspective of evaluation methodology (Weiss, 1998a) and organizational learning

(Argyris, 1999; Argyris & Schön, 1978). Theoretical inferences to evaluation methodology are strongly prominent in the data. The in-house versus outside debate is seen as important by considering trust in outside staff and their reputation, while also securing objective research. Dissemination strategies may also be crucial in order to facilitate involvement in evaluations and openness towards its findings. Most importantly, professionals care about quality and it is therefore central to create trustworthy evaluations. Findings related to evaluation methods affect learning in different forms where negative experiences creates less openness to learning, and such below-par evaluations will only be used for single-loop learning. Staff also experience limitations of learning based on political and managerial constraints that inhibit a practical focus on learning. Professionals are not negative to learning. However, they experience that learning is not readily available in a structural perspective. Evaluations are mostly used to provide solutions that affect actions, which classifies as single-loop organizational learning. However, cogent meta-evaluations may provide more potential for double-loop learning. It is more challenging to determine inferences toward organizational learning, and its separation of learning loops based on the empirical findings. Some connections have been drawn toward single-loop learning and some have shown potential for double-loop learning. Yet, the collected data is not aimed enough towards determining classifications of organizational learning in Norad. However, organizational learning is still vital to facilitate and more research will be needed to provide insight on this.

The research objectives specified depth to the research question by presenting different aspects of learning from evaluations in an organizational perspective. The thematic analysis provided the most comprehensive answers to objective 1, 2, and 4 which means that this also is most important for the participants. Attitudes toward usefulness of evaluations is connected to how the evaluation process is viewed. Evaluations are seen as less useful if one disagrees with its process or recommendations. Evaluation methodology affect area and frequency of use by pointing to single-loop learning as the outcome of decentralized evaluations and potential for double-loop learning from meta-evaluations. The centralized thematic evaluations also carry potential for single and double-loop learning, where this has been seen to be provided together with dissemination strategies. By analyzing the third and fourth objective, it is found that evaluations may play a role in organizational changes through

pointing to where efforts are needed to increase learning. However, evaluations are not seen as necessary to start changes. The discussion also finds that organizational factors in Norad point toward a focus on doing, and not learning. So far, the emphasis on learning from evaluations is mostly experienced as something spoken from management and political leadership.

On the first hand, learning from evaluations should be facilitated more systematically and culturally in Norad. Evaluations should be incorporated thoroughly into all areas of work through systematic processes that ensure utilization, so that such knowledge can be utilized across the system in decision-making processes. Such systematic processes could take the form of measures that spread evaluations and make them necessary to use in their work. In addition to this, staff need to feel as though they have the necessary space and tactics for utilizing evaluations in their work through a cultural shift. On the other hand, evaluations should also be conducted with high levels of quality so that staff will want to learn from them and so that evaluations may gain a higher reputation. Norad is a diverse organization with great responsibility to manage development funding and securing knowledge and learning of development cooperation. Based on the findings from this thesis, such recommendation will also be relevant for similar contextual organizations seeking to optimize learning from evaluations across staff.

6.2 Recommendations for Future Research

The most interesting factor for re-analysis on this topic is to study the effect of the department for knowledge in Norad. This department may affect organizational learning from evaluations once it has gained larger foothold in the organization. This thesis could not account for perceived effects from the department for knowledge, due to it being rather new and has not yet started many of the processes and systems it seeks to implement. It would be highly interesting to follow this department in a future study, because many were hopeful this would make learning more available.

This topic would also have been extremely interesting from many different theoretical viewpoints. Everything that has been left out of the theoretical synthesis could constitute new

research. Especially Caplan´s (1979) theory on the two communities seen in line with evaluators and decision makers in development cooperation. Which would give valuable insight into the dynamic relationship between researchers and policymakers to further the understanding of evaluation utilization and limiting this gap.

6.3 Limitations of this Thesis

This thesis was conducted on limited time over one semester. This created limitations in the time possible to give to recruiting participants, interviews and data analysis. A more in-depth answer to the multifaceted research question and objectives would have been possible if there were more time to give to such processes.

A prominent limitation of this thesis is the lack of variety and number of participants that were interviewed. I was only able to conduct interviews with eight participants in total, including staff in different positions within Norad either currently or previously, and one researcher. The fact that the participants have experience from different departments means that the thesis has diversity in answers and reflections. However, it also means that nothing can be concluded to one specific department or area of work. It was not possible to recruit participants from political management in this thesis, which constitutes a missing point of view that would have been extremely valuable to include for a larger overview of the Norwegian development administration.

The total number of participants and the methodological approach also mean that the findings cannot be generalized to the organization as a whole. The interviews constitute a representation of the opinions of a few voices. The participants included in the thesis were experienced as independent critical thinkers on the topic of evaluations and learning, who spoke freely of their beliefs. The participants mostly had long experience within the field of development cooperation and evaluations and much to say in this regard. Nonetheless, if other participants were interviewed, the findings may have differed because it only represents eight voices. Based on the empirical findings, it is only possible to conclude according to theory on the importance of methods for evaluations in relationship to opportunities for organizational learning in the studied context. The analysis was based the choosen theory,

which means that the findings and conclusion can only be seen as dependent on this, excluding other perspectives on the topic. The literature review and chosen theory together provided much support in shaping this research design, which also means that other inclusion or exclusion criteria could have resulted in different findings.

References

Argyris, C. (1999). *On organizational learning* (2nd ed.). Blackwell Business.

Argyris, C., & Schön, D. A. (1978). *Organizational learning : a theory of action perspective*. Addison-Wesley.

Armytage, L. (2011). Evaluating aid: An adolescent domain of practice. *Evaluation (London, England. 1995)*, *17*(3), 261-276. https://doi.org/10.1177/1356389011410518

Bakken, M. M. (2022). SIW4500 Exam Paper Autumn 2022 [Unpublished]. *OsloMet-Storbyuniversitetet*.

Berg, B. L., & Lune, H. (2017). *Qualitative Research Methods for the Social Sciences*. Pearson.

Booth, A. (2016). Searching for qualitative research for inclusion in systematic reviews: A structured methodological review. *Syst Rev*, *5*(1), 74-74. https://doi.org/10.1186/s13643-016-0249-x

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Briggs, C. L. (2003). Interviewing, Power/Knowledge, and Social Inequality In J. A. Holstein & J. F. Gubrium (Eds.), *Inside Interviewing: New Lenses, New Concerns* (pp. 495-507). Sage Publications, Inc.

Brinkmann, S. (2020). Unstructured and Semistructured Interviewing. In P. Leavy (Ed.), *The Oxford Handbook of Qualitative Research* (pp. 424–456). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190847388.013.22

Bryman, A. (2016). *Social Research Methods* (5th ed.). Oxford University Press.

Cairney, P. (2013). Standing on the Shoulders of Giants: How Do We Combine the Insights of Multiple Theories in Public Policy Studies? *Policy studies journal*, *41*(1), 1-21. https://doi.org/https://doi-org.ezproxy.oslomet.no/10.1111/psj.12000

Caplan, N. (1979). The Two-Communities Theory and Knowledge Utilization. *The American behavioral scientist*, *22*(3), 459-470. https://doi.org/10.1177/000276427902200308

Chapman, N., Lloyd, R., Villanger, E., & Gleed, G. (2017). *The Quality of Reviews and Decentralised Evaluations in Norwegian Development Cooperation* (1). (Evaluation report, Issue. Norad. https://www.norad.no/contentassets/41a0207f132c4e69a338791bac88dda7/the-

quality-of-reviews-and-decentralised-evaluations-in-norwegian-development-cooperation.pdf

Charmaz, K. (2001). Handbook of Interview Research. In. SAGE Publications, Inc. https://doi.org/10.4135/9781412973588

Durkheim, É., & Catlin, G. E. G. (1964). *The rules of sociological method* (8th ed.). Free Press of Glencoe.

Dusya, V., Crossan, M., & Apaydin, M. (2011). A Framework for Integrating Organizational Learning, Knowledge Capabilities, and Absorptive capacity. In M. Easterby-Smith & M. A. Lyles (Eds.), *Handbook of Organizational Learning and Knowledge Management* (2nd ed., pp. 153-181). John Wiley & Sons, Incorporated.

Eidhammer, A. (2007). Mytar om bistand. *Nytt norsk tidsskrift*, *24*(2), 169-176. https://doi.org/10.18261/ISSN1504-3053-2007-02-05

Emerson, R. W. (2015). Convenience Sampling, Random Sampling, and Snowball Sampling: How Does Sampling Affect the Validity of Research? *Journal of visual impairment & blindness*, *109*(2), 164-168. https://doi.org/10.1177/0145482X1510900215

Fixsen, D. L., Blase, K. A., Naoom, S. F., & Wallace, F. (2009). Core Implementation Components. *Research on social work practice*, *19*(5), 531-540. https://doi.org/10.1177/1049731509335549

Forss, K., Cracknell, B., & Samset, K. (1994). CAN EVALUATION HELP AN ORGANIZATION TO LEARN. *Evaluation Review*, *18*(5), 574-591. https://doi.org/10.1177/0193841x9401800503

Gubrium, E., & Koro-Ljungberg, M. (2005). Contending With Border Making in the Social Constructionist Interview. *Qualitative Inquiry*, *11*(5), 689-715. https://doi.org/10.1177/1077800405278776

Gaarder, M. M., & Bartsch, U. (2014). Who cares about development outcomes? Market failures and the role of the evaluation function. *Journal of Development Effectiveness*, *6*(4), 361-377. https://doi.org/10.1080/19439342.2014.966742

Gaarder, M. M., & Bartsch, U. (2015). Creating a market for outcomes: shopping for solutions. *Journal of Development Effectiveness*, *7*(3), 304-316. https://doi.org/10.1080/19439342.2015.1068357

Johannessen, L. E. F., Rafoss, T. W., & Rasmussen, E. B. (2018). *Hvordan bruke teori? : nyttige verktøy i kvalitativ analyse*. Universitetsforlaget.

Johnson, H., & Wilson, G. (2009). *Learning for development*. Zed Books.

Kvale, S., Brinkmann, S., Anderssen, T. M., & Rygge, J. (2015). *Det kvalitative forskningsintervju* (3rd ed.). Gyldendal akademisk.

Leviton, L. C., & Hughes, E. F. X. (1981). Research On the Utilization of Evaluations: A Review and Synthesis. *Evaluation Review*, *5*(4), 525-548. https://doi.org/10.1177/0193841X8100500405

Lindkvist, I., & Dixon, V. (2014). To 'feel good', or to 'do good'? Why we need institutional changes to ensure a results focus in Norwegian development assistance. *Journal of Development Effectiveness*, *6*(4), 350-360. https://doi.org/10.1080/19439342.2014.975423

Lipshitz, R. (2000). Chic, mystique, and misconception: Argyris and Schön and the rhetoric of organizational learning. *The Journal of Applied Behavioral Science*, *36*(4), 456-473.

Lloyd, R., Poate, D., & Villanger, E. (2014). Results measurement and evaluability: a comparative analysis. *Journal of Development Effectiveness*, *6*(4), 378-391. https://doi.org/10.1080/19439342.2014.966455

Lopez-Lee, D. (1982). In-house versus out-house evaluations: Administrative considerations. *Computers, environment and urban systems*, *7*(3), 163-172. https://doi.org/10.1016/0198-9715(82)90003-5

Loseke, D. R. (2017). *Methodological thinking : basic principles of social research design* (2nd ed.). SAGE Publications.

Lune, H., & Berg, B. L. (2017). *Qualitative research methods for the social sciences* (9th ed.). Pearson.

Lynn, L. E. (Ed.). (1978). *Knowledge and Policy: The Uncertain Connection*. The National Academies Press. https://doi.org/doi:10.17226/19941.

Magnusson, E., & Marecek, J. (2015). *Doing interview-based qualitative research : a learner's guide*. Cambridge University Press. https://doi.org/https://doi-org.ezproxy.oslomet.no/10.1017/CBO9781107449893

Mathison, S. (2005). *Encyclopedia of evaluation*. SAGE.

Mayer, R. E. (2012). Cognitive Learning. In N. M. Seel (Ed.), *Encyclopedia of the Sciences of Learning* (pp. 594-596). Springer US. https://doi.org/10.1007/978-1-4419-1428-6_390

Norad. (2014). *Can We Demonstrate the Difference that Norwegian Aid Makes? Evaluation of results measurement and how this can be improved* (1). https://www.norad.no/globalassets/import-2162015-80434-am/www.norad.no-ny/filarkiv/evalueringsavdelingens-filer/can-we-demonstrate-the-difference-that-norwegian-aid-makes.pdf

Norad. (2021). *Norad's strategy towards 2030* [Strategies]. https://www.norad.no/globalassets/publikasjoner/publikasjoner-2021/norads-strategy-towards-2030.pdf

Norad. (2022a). *Langsiktighet i en verden i endring. Årsrapport 2021/2022*. https://www.norad.no/globalassets/publikasjoner/publikasjoner-2022/evalueringer/arsrapport-2021-22_avdeling-for-evaluering.pdf

Norad. (2022b). *NORADS ÅRSRAPPORT 2021*. Norad. https://www.norad.no/globalassets/publikasjoner/publikasjoner-2022/norads-arsrapport-2021---15032022-med-forside.pdf

Norad. (n.d.-a). *Gjennomføring av evalueringer*. Norad. Retrieved February 02 from https://www.norad.no/evaluering/om-evaluering/gjennomforing-av-evalueringer/

Norad. (n.d.-b). *Slik er norsk bistand organisert*. Norad. Retrieved January 20 from https://www.norad.no/om-bistand/slik-er-norsk-bistand-organisert/

OECD. (2022). Norway. In *Development Co-operation Profiles*. OECD Publishing. https://doi.org/doi:https://doi.org/10.1787/aaf0304f-en

Patton, M. Q. (2008). *Utilization-focused evaluation*. Sage publications.

Prop. 1 S Gul bok (2022–2023). *For budsjettåret 2023 — Statsbudsjettet (Gul bok)*. Retrieved from https://www.regjeringen.no/contentassets/5f394e4fab7348139dc8c55156a28b81/no/pdfs/prp202220230001guldddpdfs.pdf

Reinertsen, H., Bjorkdahl, K., & McNeill, D. (2022). Accountability versus learning in aid evaluation: A practice-oriented exploration of persistent dilemmas. *Evaluation*, *28*(3), 356-378. https://doi.org/10.1177/13563890221100848

Risjord, M. W. (2014). *Philosophy of social science : a contemporary introduction*. Routledge.

Samset, K. (2015). Reell eller rituell evaluering? *Stat & styring*, *25*(3), 6-9.
https://doi.org/10.18261/ISSN0809-750X-2015-03-03

Samset, K. F., Forss, K., Hauglin, O., & Scanteam, I. (1993). *Learning from experience : a study of the feedback from evaluations and reviews in Norwegian bilateral aid*. Royal Ministry of Foreign Affairs.

Ternström, I., Baker, J., Dahlgren, S., Lithman, E., Rehman, A., & Bhattacharjee, A. (2020). *Quality Assessment of Decentralised Evaluations in Norwegian Development Cooperation (2018-2019)*.
https://www.norad.no/contentassets/e7901444a02a4beb97462a50773e535c/6.20-quality-assessment-of-decentralised-evaluations-in-norwegian-development-cooperation-2018-2019.pdf

Ternström, I., Baker, J., Dahlgren, S., Lithman, E., Rehman, A., & Bhattacharjee, A. (2021). *Annual Quality Assessment of Reviews in Norwegian Development Cooperation (2019–2020)*.
https://www.norad.no/contentassets/977919df84ba4ed7824ec7ca5f1cc71d/3.21_annual_quality_assessment_of_reviews_in_norwegian_development_cooperation_2019-2020.pdf

Tjønneland, E. (2018). Kvalitet og bruk av evaluering i norsk utviklingshjelp. *Stat & styring*(3), 32-33.

Traianou, A. (2020). The Centrality of Ethics in Qualitative Practice. In P. Leavy (Ed.), *The Oxford Handbook of Qualitative Research* (pp. 86–110). Oxford University Press.
https://doi.org/10.1093/oxfordhb/9780190847388.013.12

Villanger, E., & Jerve, A. M. (2009). Assessing aid impact: a review of Norwegian evaluation practice. *Journal of Development Effectiveness*, *1*(2), 171-194.
https://doi.org/10.1080/19439340902903229

Villanger, E., Lloyd, R., Poate, D., & Ali, M. (2016). Why can we not demonstrate the difference that Norwegian aid makes? An evaluation of the results measurement system and practice. *Journal of Development Effectiveness*, *8*(2), 171-195.
https://doi.org/10.1080/19439342.2015.1096297

Weiss, C. H. (1978). Improving the Linkage Between Social Research and Public Policy In L. E. Lynn (Ed.), *Knowledge and Policy: The Uncertain Connection* (Vol. 5, pp. 23-81). The National Academies Press.

Weiss, C. H. (1979). The Many Meanings of Research Utilization. *Public administration review*, *39*(5), 426-431. https://doi.org/10.2307/3109916

Weiss, C. H. (1998a). *Evaluation : methods for studying programs and policies* (2nd ed.). Prentice Hall.

Weiss, C. H. (1998b). Have we learned anything new about the use of evaluation? *The American journal of evaluation*, *19*(1), 21-33. https://doi.org/10.1016/S1098-2140(99)80178-7

Wildavsky, A. (1972). The Self-Evaluating Organization. *Public administration review*, *32*(5), 509-520. https://doi.org/10.2307/975158

World Bank Group. (2017). *Governance and the law*. World Bank Group.

Xiao, Y., & Watson, M. (2019). Guidance on Conducting a Systematic Literature Review. *Journal of planning education and research*, *39*(1), 93-112. https://doi.org/10.1177/0739456X17723971

Appendix 1: PRISMA Flow Chart
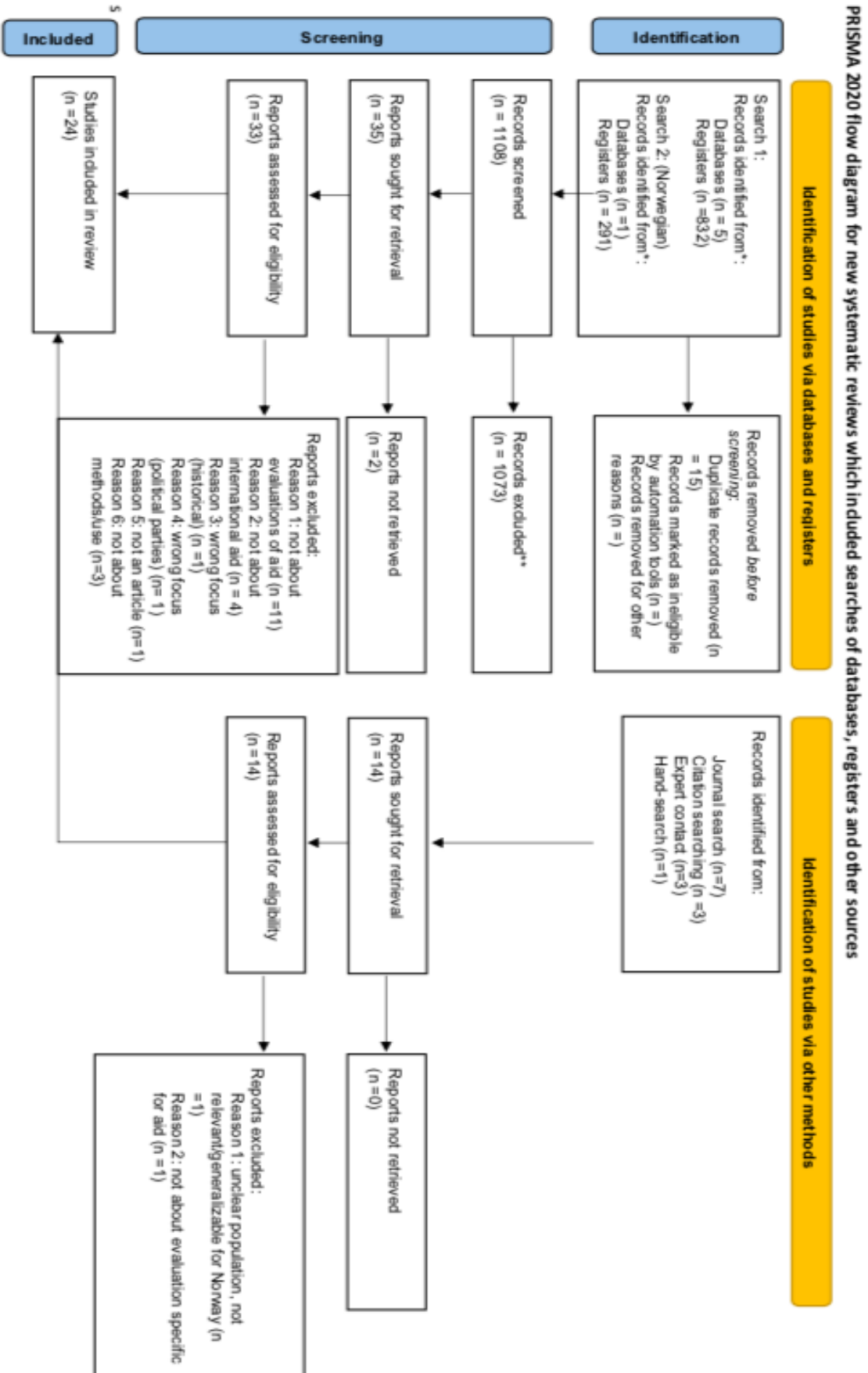
PRISMA 2020 flow diagram for new systematic reviews which included searches of databases, registers and other sources

Identification of studies via databases and registers

Identification

Search 1:
Records identified from*:
Databases (n = 5)
Registers (n =832)

Search 2: (Norwegian)
Records identified from*:
Databases (n =1)
Registers (n = 291)

Records removed before screening:
Duplicate records removed (n = 15)
Records marked as ineligible by automation tools (n =)
Records removed for other reasons (n =)

Screening

Records screened
(n = 1108)

Records excluded**
(n = 1073)

Reports sought for retrieval
(n =35)

Reports not retrieved
(n =2)

Reports assessed for eligibility
(n =33)

Reports excluded:
Reason 1: not about evaluations of aid (n =11)
Reason 2: not about international aid (n = 4)
Reason 3: wrong focus (historical) (n =1)
Reason 4: wrong focus (political parties) (n= 1)
Reason 5: not an article (n=1)
Reason 6: not about methods/use (n=3)

Included

Studies included in review
(n = 24)

Identification of studies via other methods

Records identified from:
Journal search (n =7)
Citation searching (n =3)
Expert contact (n =3)
Hand-search (n=1)

Reports sought for retrieval
(n =14)

Reports not retrieved
(n=0)

Reports assessed for eligibility
(n =14)

Reports excluded:
Reason 1: unclear population, not relevant/generalizable for Norway (n =1)
Reason 2: not about evaluation specific for aid (n =1)

Appendix 2: Explanation of PRISMA Flow Chart and Search Strategy

The literature search strategy for the literature review was put together after a meeting with a librarian. This meeting was requested due to difficulties finding the right search terms for relevant literature. Based on recommendations from the librarian and personal experiences, the decision was made to broaden the search terms and included databases. This strategy included a search in both English and Norwegian, as well as conducting additional searches outside the included databases. The first search was done in the databases Academic Search Ultimate, Econlit, Socindex, Web of Science, and Assia. Using the following search terms: (*Norway OR Norwegian) AND ("Development aid" OR "development assistance" OR "international economic assistance" OR "foreign aid" OR "international aid") AND (learn* OR adapt* OR modify* OR alter* OR change* OR lessons* OR improve*) AND (evaluation* OR "evaluation methods" OR "evaluation process").* This search gave 832 results in total.

I then conducted the same search translated to Norwegian in Oria, which gave 291 results. All articles were exported to a new library in EndNote, before applying the function to extract all duplicates. After going through the results to manually also remove duplicates, 15 articles were removed that had been exported in duplicates, leaving 1108 articles left to screen. The decision to conduct a broad search was made to ensure that all relevant literature was included. Based on the initial searches which did not give high results, it was seen as a better strategy to rather screen out many articles later on. To remove all irrelevant articles as efficiently as possible, I went through the list alphabetically to remove what I was sure was not related to this topic based on titles. The potentially relevant articles were screened through reading the abstract. This left 35 articles to screen more thoroughly, where two were not possible to retrieve. The inclusion criteria at this stage was as follows: study done in Norway or relevancy to Norwegian aid administration, study on evaluations or evaluation methods for international aid. Some articles were therefore removed during screening because they were not about international aid or dealing with questions of evaluations of international aid. One was removed because it was not an article, and two were removed because the focus of the article was not relevant. Lastly, three articles were removed because they did not go into usage or methods for evaluation of aid.

The additional search started with a site search in "Journal of Development Effectiveness", with the search terms "Norway OR Norwegian". This gave 27 results, where seven were a fit to the inclusion criteria and was exported to EndNote. Following a hand search in "Stat & Styring" going through every publication dating five years back from 2022 where they had four publications every year, one relevant article was extracted. Two articles were included from expert contact through email. Lastly, a citation search was conducted on a few central articles resulting in three exported articles. The additional search resulted in 13 studies for further screening. One was removed because the evaluation focus was not relevant for international aid, and another was removed because it was not relevant for the Norwegian context. Adding the results from both searches, 24 studies were included in the review. Many of these have been used in this master's thesis and several have been excluded in the writing process.

Appendix 3: NSD Approval

◀◑ Sikt

Vurdering av behandling av personopplysninger

Referansenummer	**Vurderingstype**	**Dato**
994589	Standard	16.11.2022

Prosjekttittel
Professional Attitudes on the Focus of Learning from Evaluations

Behandlingsansvarlig institusjon
OsloMet – storbyuniversitetet / Fakultet for samfunnsvitenskap / Institutt for sosialfag

Prosjektansvarlig
Simon Innvær

Student
Maria Bakken

Prosjektperiode
01.12.2022 - 10.05.2024

Kategorier personopplysninger
Alminnelige

Lovlig grunnlag
Samtykke (Personvernforordningen art. 6 nr. 1 bokstav a)

Behandlingen av personopplysningene er lovlig så fremt den gjennomføres som oppgitt i meldeskjemaet. Det lovlige grunnlaget gjelder til 10.05.2024.

Meldeskjema ☑

Kommentar

ABOUT OUR ASSESSMENT
Data Protection Services has an agreement with the institution where you are carrying out research or studying. As part of this agreement, we provide guidance so that the processing of personal data in your project is lawful and complies with data protection legislation.

We have now assessed the planned processing of personal data. Our assessment is that the processing is lawful, so long as it is carried out as described in the Notification Form with dialogue and attachments.

IMPORTANT INFORMATION
You must store, send and secure the collected data in accordance with your institution's guidelines. This means that you must use online survey, cloud storage, and video conferencing providers (and the like) that your institution has an agreement with. We provide general advice on this, but it is your institution's own guidelines for information security that apply.

TYPE OF DATA AND DURATION
The project will be processing general categories of personal data until the date documented in the Notification form.

LEGAL BASIS
The project will gain consent from data subjects to process their personal data. We find that consent will meet the necessary requirements under art. 4 (11) and 7, in that it will be a freely given, specific, informed and unambiguous statement or action, which will be documented and can be withdrawn.

The legal basis for processing general categories of personal data is therefore consent given by the data subject, cf. the General Data Protection Regulation art. 6.1 a).

PRINCIPLES RELATING TO PROCESSING PERSONAL DATA
We find that the planned processing of personal data will be in accordance with the principles under the General Data Protection Regulation regarding:
• lawfulness, fairness and transparency (art. 5.1 a), in that data subjects will receive sufficient information about the processing and will give their consent
• purpose limitation (art. 5.1 b), in that personal data will be collected for specified, explicit and legitimate purposes, and will not be processed for new, incompatible purposes
• data minimisation (art. 5.1 c), in that only personal data which are adequate, relevant and necessary for the purpose of the project will be processed
• storage limitation (art. 5.1 e), in that personal data will not be stored for longer than is necessary to fulfil the project's purpose

THE RIGHTS OF DATA SUBJECTS

As long as the data subjects can be identified in the data material, they will have the following rights: access (art. 15), rectification (art. 16), erasure (art. 17), restriction of processing (art. 18), data portability (art. 20).

We find that the information that will be given to data subjects about the processing of their personal data will meet the legal requirements for form and content, cf. art. 12.1 and art. 13.

We remind you that if a data subject contacts you about their rights, the data controller has a duty to reply within a month.

FOLLOW YOUR INSTITUTION'S GUIDELINES

We presuppose that the project will meet the requirements of accuracy (art. 5.1 d), integrity and confidentiality (art. 5.1 f) and security (art. 32) when processing personal data.

To ensure that these requirements are met you must follow your institution's internal guidelines and/or consult with your institution (i.e. the institution responsible for the project).

NOTIFY CHANGES

If you intend to make changes to the processing of personal data in this project it may be necessary to notify us. This is done by updating the Notification Form. On our website we explain which changes must be notified: https://www.nsd.no/en/data-protection-services/notification-form-for-personal-data/notify-changes-in-the-notification-form

Wait until you receive an answer from us before you carry out the changes.

FOLLOW-UP OF THE PROJECT

We will follow up the progress of the project at the planned end date in order to determine whether the processing of personal data has been concluded.

Good luck with the project!

Sikt

Vurdering av behandling av personopplysninger

Referansenummer	**Vurderingstype**	**Dato**
994589	Standard	03.02.2023

Prosjekttittel
Professional Attitudes on the Focus of Learning from Evaluations

Behandlingsansvarlig institusjon
OsloMet – storbyuniversitetet / Fakultet for samfunnsvitenskap / Institutt for sosialfag

Prosjektansvarlig
Simon Innvær

Student
Maria Bakken

Prosjektperiode
01.12.2022 - 10.05.2024

Kategorier personopplysninger
Alminnelige

Lovlig grunnlag
Samtykke (Personvernforordningen art. 6 nr. 1 bokstav a)

Behandlingen av personopplysningene er lovlig så fremt den gjennomføres som oppgitt i meldeskjemaet. Det lovlige grunnlaget gjelder til 10.05.2024.

Meldeskjema ↗

Kommentar
Data Protection Services has assessed the change registered on 03.02.2023.

We find that the processing of personal data in this project is lawful and complies with data protection legislation, so long as it is carried out as described in the Notification Form with dialogue and attachments.

The project has added Zoom as data processor.
When using a data processor (questionnaire provider, cloud storage, video call etc.), the processing must meet the requirements for the use of a data processor, cf. art. 28 and art. 29. Use suppliers with whom your institution has an agreement.

FOLLOW-UP OF THE PROJECT
We will follow up the progress of the project underway (every other year) and at the planned end date in order to determine whether the processing of personal data has been concluded/is being carried out in accordance with what is documented.

Contact person at Data Protection Services:
Henriette S. Munthe-Kaas

Good luck with the project!

Appendix 5: Interview Guide

Bakgrunnsinformasjon:

1. kan du fortelle meg litt om den typiske gangen i en evaluering av et bistandsprosjekt, fra initiativfasen til sluttevalueringen?

    a. Hva vektlegges mest når resultatene av en evaluering skal vurderes?

2. hvordan vil du beskrive din interesse/ditt forhold til evalueringer av bistandsprogrammer?

Evaluerings design for læring

1. hva er dine tanker om å bruke evalueringer for å vise ansvar ovenfor et program?

    a. Ansvar ovenfor befolking hjemme/mottaker/ andre som jobber med prosjektet?

2. utviklingshjelp har ofte komplekse mål med mange viktige områder for utvikling, hva tenker du rundt dilemmaet med å evaluere dette?

    a. Hvordan kan man best observere at mål blir oppnådd?

3. evalueringer er ofte desentraliserte, hvordan tror du dette kan påvirke bruken og læringsutbyttet fra evalueringer? (med tanke på at de ikke er involvert i selve bistanden?)

    a. hva tror du er den beste måten å jobbe med evalueringer? – med tanke på involvering/ internt-eksternt

Tanker om viktighet av evalueringer

1. hva er din generelle mening om viktigheten av evalueringer av bistand?

    a. Om ikke viktig, hva er det viktigste innen bistand for læring og hvordan oppnår man det?

    b. Om viktig, referer til studie

2. hvordan kan evalueringer bidra til endring av pågående eller framtidige programmer?

    a. Hvordan ser forvaltningen av kunnskapen ut i slike situasjoner?

3. kan du beskrive viktigheten av å inkludere «lessons learned» i evalueringer?

4. kan du tenke på en spesifikk evaluering som hadde mye betydning/læring? Hvorfor var den viktig?

Forskjellige måter å lære fra evalueringer

1. hvordan føler du at å være involvert i en evaluering selv kan ha påvirkning på lærdommene du/ andre tar med seg videre?

    a. På hvilke måter kan man være involvert/til hvilken grad?

2. kan du si litt om hvordan evalueringer kan brukes som et verktøy på tvers av bistandsprogrammer?

    a. På hvilke måter kan de brukes som støtte i forskjellige arbeidsoppgaver i forskjellige roller?

3. på hvilke måter tror du det er mulig å inkorporere lærdommer fra evalueringer inn i Norad sin kunnskap-struktur? hvordan ser denne strukturen ut? (utviklet med mye kunnskap og mange kilder?)

    a. Hvordan kan de brukes for å korrigere feil eller effektivisere noe?

    b. Hvordan kan informasjonen man får overskride selve programmet som blir evaluert, kan de brukes for å utvikle noe nytt eller omvelte det eksisterende?

    c. hvordan tror du at evalueringer kan spille en rolle i å bryte normer?

Erfaringer med strukturer/miljø for læring

1. hvordan vil du beskrive fokuset på å lære fra og bruke evalueringer?

    a. Hvor kommer dette fokuset fra?

2. kan du si litt om dine tanker om forholdet mellom organisasjonsstruktur/kultur og rollen til evalueringer?

    a. Hvordan føler du at ansatte/profesjonelle har incentiver til å bruke evalueringer?

3. kan du beskrive tiden og ressursene brukt rundt evalueringer av programmer?

Avslutning:

1. er det noe du ønsker å legge til?

2. mulighet for fremtidig kontakt

3. tusen takk

Appendix 6: Participant Consent Form

Are you interested in taking part in the research project:
"Professional Attitudes on the Focus on Learning from Evaluations"?

Purpose of the project

You are invited to participate in a research project where the main purpose is to investigate how professionals in the field of evaluations and foreign aid think about the focus of learning from evaluations. And how they believe evaluations may be important in their work. This master´s thesis will be based on the following research question: What are the attitudes held by professionals in the evaluation field on the focus on learning and lessons learned in impact evaluations, and how may this be related to the organizational environment?

Which institution is responsible for the research project?

Oslo Metropolitan University is responsible for the project (data controller).

Why are you being asked to participate?

The sample of participants have been selected based on current or previous professional experience with evaluations in Norad. And because they have information and experience on the topic and research question. The sample has been selected based on conversations with others, and through information on websites.

What does participation involve for you?

If you choose to participate in the projects, this will involve that you take part in an interview. It will take approximately one hour. The interview includes questions about your thoughts about the importance of evaluations of foreign aid, how they may be conducted in a way that allows for learning and the working environment surrounding evaluations. Your answers will be recorded with sound.

Participation is voluntary

Participation in the project is voluntary. If you chose to participate, you can withdraw your consent at any time without giving a reason. All information about you will then be made anonymous. There will be no negative consequences for you if you chose not to participate or later decide to withdraw.

Your personal privacy – how we will store and use your personal data

We will only use your personal data for the purpose(s) specified here and we will process your personal data in accordance with data protection legislation (the GDPR).

- Maria Bakken (student) and Simon Innvær (supervisor) will have acces to the personal data.
- I will replace your name and contact details with a code.
- Participants will not be recognizable in any publications
- The data will be stored on a research server

What will happen to your personal data at the end of the research project?

The planned end date of the project is 10.05.2024.

I will delete all sound recordings and all personal identification correspondence. The non-identifying transcriptions will be saved for the possibility of further research.

Your rights

So long as you can be identified in the collected data, you have the right to:
- access the personal data that is being processed about you
- request that your personal data is deleted
- request that incorrect personal data about you is corrected/rectified
- receive a copy of your personal data (data portability), and
- send a complaint to the Norwegian Data Protection Authority regarding the processing of your personal data

What gives us the right to process your personal data?

We will process your personal data based on your consent.

Based on an agreement with Oslo Metropolitan University, Data Protection Services has assessed that the processing of personal data in this project meets requirements in data protection legislation.

Where can I find out more?

If you have questions about the project, or want to exercise your rights, contact:
- Oslo Metropolitan University via Simon Innvær: simon.innvar@oslomet.no
- Our Data Protection Officer: Ingrid Jacobsen

If you have questions about how data protection has been assessed in this project, contact:
- Data Protection Services, by email: (personverntjenester@sikt.no) or by telephone: +47 53 21 15 00.

Yours sincerely,

(Researcher/supervisor)   (Student)

Consent form

I have received and understood information about the project "Professional attitudes on the Focus on Learning from Evaluations" and have been given the opportunity to ask questions. I give consent:

☐   to participate in an interview

I give consent for my personal data to be processed until the end of the project.

(Signed by participant, date)