Fall 1-4-2024

# Unpacking Challenges and Reinforcing Evaluations in Bilateral and Multilateral Development Organizations in Egypt

fatema elghonemy
fatemaelghonemy@aucegypt.edu

**The American University in Cairo**

**School of Global Affairs and Public Policy**

# Unpacking Challenges and Reinforcing Evaluations in Bilateral and Multilateral Development Organizations in Egypt

A Thesis Submitted to the

Public Policy and Administration Department

in partial fulfillment of the requirements for the degree of

Master of Public Administration

By

## Fatema Mahmoud El Ghonemy

Supervised by:

Dr. Laila El Baradei

Professor of Public Administration

**Fall 2023**

## *Table of Contents*

## *List of Acronyms*

| | |
|---|---|
| **AEA** | American Evaluation Association |
| **BPRM** | Bureau of Population, Refugees, and Migration |
| **BMZ** | Federal Ministry of Economic Cooperation and Development (Germany) |
| **DAC** | Development Assistance Committee |
| **EEDS** | Enhancement of the Egyptian Dual System Project |
| **EGRA** | Early Grade Reading Assessment |
| **EPP** | Employment Promotion Project |
| **EU** | European Union |
| **FDA** | French Development Agency |
| **FGDs** | Focus Group Discussions |
| **GIZ** | Deutsche Gesellschaft für Internationale Zusammenarbeit |
| **GT** | Grounded Theory |
| **IFPRI** | International Food Policy Research Institute |
| **IRB** | Institutional Review Board |
| **J-PAL** | Abdul Latif Jameel Poverty Action Lab |
| **JICA** | Japan International Cooperation Agency |
| **KII** | Key Informant Interviews |
| **MDGs** | Millennium Development Goals |
| **M&E** | Monitoring and Evaluation |
| **MPED** | Ministry of Planning and Economic Development |
| **MoETE** | Ministry of Education and Technical Education |
| **ODA** | Official Development Assistance |
| **OECD** | Organization for Economic Co-operation and Development |

| **RBM** | Results-Based Management |
|---|---|
| **RCTs** | Randomized Controlled Trials |
| **RISE** | Realizing Innovation through Social Entrepreneurship |
| **SDGs** | Sustainable Development Goals |
| **TCTI** | Technical Support for the Comprehensive Technical Education Initiative with Egypt |
| **UN** | United Nations |
| **UNDP** | United Nations Development Programme |
| **UNICEF** | United Nations International Children's Emergency Fund |
| **UNPDF** | United Nations Partnership Development Framework |
| **USAID** | United States Agency for International Development |
| **WBG** | World Bank Group |

## List of Figures

## List of Tables

## *Acknowledgments*

The American University in Cairo
School of Global Affairs and Public Policy
Department of Public Policy and Administration

# Unpacking Challenges and Reinforcing Evaluations in Bilateral and Multilateral Development Organizations in Egypt

By
Fatema Mahmoud El Ghonemy
Supervised by
Dr. Laila El Baradei

## *Abstract*

The research examines evaluations conducted in bilateral and multilateral development organizations considering programs that are funded by governments or international organizations. The Egyptian Government partners with development partners to receive technical assistance in various fields, such as industrial development, technical education, environmental protection, health improvement and refugees' empowerment. The thesis explores the interlinkage between the development goals and the results framework to achieve sustainable development goals (SDGs). It then goes on to examine the evolution of the function of evaluation and its historical development into becoming a standalone field that serves the decision making, extension and expansion of interventions. As the frequency of international development interventions increases, so does the need to ensure that they are effective, efficient, impactful, sustainable, relevant and coherent - in accordance with what is known as OECD - DAC Criteria Framework to ensure aid effectiveness - and this can be accomplished through evaluations. The thesis used qualitative methodology by conducting 19 KIIs with practitioners, researchers, agencies' staff members and decision makers as well as government officials. It is important to understand that evaluating development partnership projects in Egypt has its own challenges that hinder the betterment of international development interventions. Understanding the significance of evaluations for projects is the main objective of the thesis as well as highlighting the missed opportunities of not conducting comprehensive evaluations. The findings showed the difference between tick-the-box evaluations and rigorous evaluations while keeping in mind the different factors that should be considered in order to call an evaluation "strong" and "comprehensive". It sheds light on the different stakeholders and their varying objectives while exploring the possible ways of integrating them into an evaluation participatory approach to maximize evaluation utilization. The recommendations generated by the thesis tackle each challenge by calling for a unified dictionary for staff working in evaluation in Egypt, strengthening ownership for public partners, using multi-layer and independent evaluation institutions and developing capabilities of evaluation stakeholders.

## *Chapter One: Introduction*

International development co-operations were established right after World War II in the mid-1940s in an attempt to improve global development and tackle the many rising risks that were affecting humanity. Even though billions of dollars are spent on development assistance yearly from development cooperations implemented in developing countries, poverty rates continue to rise, the environment continues to suffer, and more calamities continue to arise (Baker, 2000; Koolwal et al., 2010; Roob, 2014). In the rapid evolution of development projects being initiated and working towards achieving the Sustainable Development Goals (SDGs) for 2030, evaluation is also taking large steps into becoming a standalone field with crucial approaches to measure the value and objectives of the different development projects and interventions implemented.

Egypt adopted an open and participatory approach in development starting from the 2000s that integrates civil society organizations, whether local or international. According to the latest numbers reported, Egypt has received on average 1 billion USD every year from 2005 onwards from development assistance as listed in OECD, being a lower-middle income country (OECD, 2023). International development assistance started in Egypt in the 1940s with the seven key development partners: the World Bank Group (WBG), European Union (EU), United States Agency for International Development (USAID), the German Agency for International Cooperation (GIZ), United Nations Development Programme (UNDP), Japan International Cooperation Agency (JICA), and the French Development Agency (FDA). Egypt received 85 billion USD between 1946 and 2020 from the U.S. Agency for International Development (USAID) (AmCham Egypt, 2021). In FY 2021, Egypt received 1.29 billion USD from the U.S.

only (Haines, 2023). The benefits derived from development projects for economic growth, poverty alleviation, health, and investment in human capital can be seen broadly. Since 1956, Egypt and Germany have worked closely together in the field of development cooperation. The ongoing portfolio of the Egyptian-German cooperation is estimated at 1.7 billion Euro (2019) in loans, financial contributions, grants and technical cooperation. Yet, there is a lack of absorptive capacity of the Egyptian government for structural reforms that addresses the root causes of development challenges rather than the symptoms; thus, enabling the country to become less dependent on aid and support from development partners. One of the first approaches taken towards strengthening monitoring and evaluation (M&E) efforts to achieve better aid effectiveness was the "Managing for Results" approach by OECD in its Egypt chapter in 2011 (OECD, 2011). The approach relied on connecting resources given to programs by donors and partners with well-defined results and assessing the progress on the projects as well as learning from the produced information to make evidence-based decisions. The Paris Declaration in 2010 aimed to enhance monitoring and evaluation frameworks on a global scale and became the first instance in Egypt in which a 'results-oriented framework' was established for development objectives (OECD, 2011). The Ministry of Planning and International Cooperation was the first government entity to oversee the process and it introduced a results-based management (RBM) approach that was implemented to monitor other ministries. However, The Ministry of International Cooperation established a special unit called PEMA 'Project Evaluation and Macroeconomic Analysis Center' that was established in 2002 (Interviewee Sixteen, 2023). In an assessment report by the World Bank in 2007 (OECD), they announced that Egypt earned a score of "B" in meeting Paris Declaration targets and emphasized the role of the results framework and

the reliability and high quality of work that was done by Egypt. However, one of the weaknesses reported was lack of available data, specifically baseline data (OECD, 2023).

## 1.1: Research Problem

Although there is no unified understanding of program evaluation, all evaluation definitions stress the importance of understanding the worthiness of interventions and whether they are helpful to beneficiaries or not. Evaluation matters in this specific time in history because it contributes to better accountability and more focused results that speak for better value for money as well as effective public good (Behn, 2003; Holzapfel, 2016; Hoefer, 2000; Liket et al., 2014). Adoption of evaluation when designing an intervention brings better value for the money that is either donated by individuals and entities or taken as part of taxes and utilized by bilateral and multilateral development agencies. Development agencies craft their own results frameworks to measure the efficacy of the monitored processes and to ensure that they utilize the most effective approach to ensure public good (Holzapfel, 2016). The frameworks used to measure results are required to include standard indicators that are then aggregated across several interventions happening at the same time between various agencies that measure the same indicators. This increases the sense of accountability and effectiveness of interventions when it comes to implementers and development partner organizations (Holzapfel, 2016).

Behn outlined the intent behind pursuing evaluations in 4 main reasons: to gain control of the organization through the implementation of rules and enforcing adherence to them; to monitor budget spending and ensure that it is conducted efficiently and in alignment with the available resources; to motivate staff members and stakeholders with information on

performance; and to learn from previous mistakes and find ways to correct the failed processes in the future (2003, p.586). This notion is largely supported by Liket's survey which he conducted to determine organizations' motivation to conduct evaluations. He found that 57% of organizations do so to ensure internal accountability and control over their services and resources; 51% wish to see their impact in reality; 42% use their evaluations for media and external communication, and 39% of the surveyed pursue evaluations to check the box of the funder requirements (2014).

Following a thorough literature review on the subject, it became clear that there is a lack in literature and resources that assess the evaluation efforts of development projects in Egypt, including their impact and evaluation methods limitations. In general, it appears that research on evaluations for international development partners is extremely scarce. Despite the amount of money spent on projects and interventions from development agencies (Haines, 2023), the results are mostly unknown to the public. If the purpose of development aid is to enable recipients and recipient countries, promote positive changes in their lives and well-being, and reduce poverty, then this raises questions about the aid community's long-standing involvement in Egypt, which spans more than 50 years (Elbaradei, 2007). It calls into question whether a greater level of national ownership and self-reliance will be attained, and when outside development assistance will eventually be unnecessary (Glewwe & Todd, 2022; Holzapfel, 2016). The need for rigorous evaluation has been noted by J-PAL in Egypt that motivated establishing the Egyptian Impact Lab in 2022. The contribution of this study is to provide a stepping stone in the importance of comprehensive evaluation as well as support further research that can critique and recommend

11

evaluation approaches and practices for development programs in Egypt, and to leverage their learning in informing future design and implementation of development projects and policies.

The primary objective of aid project evaluations is to enhance internal accountability and efficiency in development agencies. The absence of ownership and comprehension of evaluations and log frames by governmental counterparts, who hold a pivotal position in guaranteeing the effectiveness of aid, is a noteworthy concern. Both sides continue to obfuscate the political economy underpinnings of aid and development initiatives, which makes data scarce and results of evaluations reluctant to be made public. Furthermore, there is a need for more structured and organized evaluation processes because the current method of evaluating development projects is insufficiently systematic.

### 1.2: Research Focus and Questions

The study aims to draw an analytical framework that pinpoints current evaluation practices and challenges facing development partner organizations and focuses on developmental projects funded by most-well known organizations working in Egypt such as USAID, GIZ, WFP, UNICEF and UNDP. The dilemma can be summarized as follows: if evaluations are not designed and implemented effectively to fulfill specific aims, and interventions continue without modification or reiterative amendments, then resources will be wasted and beneficiaries will not be better off which would make development cooperation initiatives of little benefit. The study does not focus on local nor international NGOs, nor does it take into account government-led programs and initiatives. Rather, it focuses on evaluation as part of the M&E system and explores the relationship between aid effectiveness and evaluations in specific. In order to do so, the

conceptual framework, developed by the researcher as a curation of UN selected agencies, GIZ and USAID evaluation framework, is an attempt to pave the way for more rigorous evaluation inspired by research findings while taking into consideration the Egyptian context and the nature of bilateral and multilateral partnerships.

The literature review draws attention to the ambiguity that surrounds the concept of programme evaluation, with different interpretations ranging from internal monitoring practices to rigorous impact evaluations using strong quantitative methodologies (El Baradei et al., 2014; Liket et al., 2014; Mulligan & Sherriff, 2019). The purpose of this study is to clarify the variables that affect thorough evaluation through an analysis of the main partners in bilateral and multilateral agreements, including US, German, WFP, UNICEF and UNDP, as well as their competing goals. The study notably highlights the distinction between evaluative entities and implementing bodies (public institutions), emphasizing the need for independent evaluators for improved accountability and the ultimate objective of attaining governmental or national ownership. The research delves into different evaluation types, elucidating their expected outcomes, and explores evaluation at various levels, be it policy or organizational, to fulfill diverse goals such as providing insights, informing decisions, learning from past projects, or supporting legislative efforts by policymakers.

**Main Research Question**:

How do evaluation practices shape and influence the opportunities and challenges in donor-funded development projects and programs implemented in cooperation with the Egyptian Government?

**First Sub-Question:** What is the status quo of evaluations for development projects in Egypt in bilateral and multilateral organizations?

**Second Sub-Question:** How do prevailing evaluation approaches contribute to the effectiveness of development projects?

**Third Sub-Question:** What are the evidence-based recommendations that could be leveraged to evaluate development projects in Egypt?

The first two questions will be answered thoroughly in the findings section by tackling the main challenges and sub-challenges that are drawn from the status quo and that provide better approaches for effective evaluation. The third question will be answered through chapter seven in the recommendations and remarks section.

### 1.3: Thesis Composition

The thesis is organized into six main chapters, which focus on the following aspects:

**Chapter One** introduces the research topic and the problem addressed in this thesis. It explores the research objectives and outlines the scope of the study. Additionally, this introductory chapter delves into the methodology employed and presents how the study's findings are answering the main questions that this work aims to address.

**Chapter Two** conceptualizes the framework for the thesis, drawing insights from the literature review and primary research findings. It delineates the placement of each concept within the comprehensive framework, organized into categories: evaluation types, determinants, purposes, and the involved stakeholders in bilateral and multilateral partnerships.

**Chapter Three** illustrates a review of available literature within the research focus. It begins by tracing the historical development of evaluation, distinguishing between evaluation and

research, and underscoring the essential role of evaluations. The chapter then examines the prevalent evaluation frameworks in bilateral and multilateral development projects. Finally, it delves into the Egyptian context, elucidating the entities engaged in evaluation and their ongoing efforts in this domain.

**Chapter Four** delves into the research methodology employed for this study, explaining the justification for opting for qualitative research methods and detailing the data collection techniques. This section elaborates on the rationale behind sample selection, the study framework, the process of data analysis, ethical considerations, and outlines the limitations inherent in the study.

**Chapter Five** provides an overview of the nexus between international development and evaluation in terms of policy framework. It begins by examining the frameworks established in MDGs and SDGs for evaluating results, progressing to an exploration of the reasons behind the challenges associated with development. The chapter emphasizes the necessity for evaluation to address these issues and introduces various efforts made by the OECD in this context in Egypt.

**Chapter Six** analyzes the findings from the nineteen interviews conducted by the organizations selected for the study. The results are divided into four main themes: (1) 'Impact does not happen overnight': Raising challenges in measuring projects' impact; (2) Evaluating Significance-related Challenges; (3) Stakeholders Dynamics Challenge; and (4) Evaluation Utilization.

**Chapter Seven,** recommends to the study that answer all the challenges encountered in the analysis through strengthening stakeholder dynamics, developing capabilities, multi-layer evaluation and independent evaluation body then it concludes the final remarks about the study.

## *Chapter Two: Conceptual Framework*

Through delving into the secondary and primary sources, the conceptual framework is providing a holistic approach that gathers the contextual information about Egypt's bilateral and multilateral organizations as well as the correlation between international development and evaluation that can support the transition from conducting poor to comprehensive development for evaluations. **The main variables are evaluation, rigorous evaluation, and accountability of donors, ownership of counterparts, and independence of the evaluation agency.**

The conceptual framework is derived from literature review as well as field work findings especially, El Baradei et al., 2014; the field work findings in specific Interviewees Eight, Nine and Sixteen, 2023; Ravallion, 2009; Orr, 2018. It starts from the evaluation arrow, according to literature evaluation types include but not only "monitoring, process evaluation, cost-benefit evaluation, and impact evaluation" (Baker, 2000, P. 45). The one-point researchers confirmed and agreed on is that the process to get rigorous evaluations will not happen through one method only but by understanding that the methods complement each other.

The figure is answering the research questions, starting from what the status quo for evaluation and that is was answered by reading the intervention triangle that cannot happen except by the three main stakeholders: Development Partner Organization (Funder); Public Partner Institution (Implementer); and Evaluation body. The challenges dynamics between the three stakeholders was analyzed through theme 6.3 in the findings. The last sub-question addressed the evidence-based recommendations that are written in yellow in the triangle and

inspired calls for organizations to be more accountable about their interventions, calling public partners to increase their sense of ownership and calling evaluation bodies to be independent and not funded by the organization itself (El Baradei et al., 2014; Ravallion, 2009; Interviewee Eight, Nine and Sixteen, 2023). The main research question about challenges and opportunities is answered through the determinants written in blue in the figure that either hinder the road to comprehensive evaluations or was an opportunity to better evaluate the project. Finally, the purpose, types and criteria are answering the sub-question about the best approaches for evaluation methods. Since this part was founded that no one type of evaluation can be the best approach, but it should be chosen based on the intervention, purpose of evaluation criteria chosen (Baker, 2000; Orr, 2018).

## 2.1: Bilateral and Multilateral Partnership: The three stakeholders

The main variables are accountability of donors, ownership of counterparts, and independence of evaluation agencies. But to understand the variables, it is crucial to understand the partnership dynamics first. The partnership is composed of a development partner organization and in our case it is either the German, American or selected UN bodies and they play the funder arm, taking their money from taxpayers and paying for developing countries (ODA). The second partner is the government institute that plays the implementer arm that has different roles in the partnership starting from agreeing on the program and signing a MoU until receiving the technical assistance and resources to implement the intervention. Coming in third is the evaluator body. In Egypt, the evaluator is commissioned by the development partner organization to externally evaluate the process. The analysis brought what the literature could not

offer, mainly that development partner organizations have to work on increasing their accountability by supporting the public partner to expand their ownership of the projects and by agreeing on having an independent evaluation body conduct their evaluation for better transparency and to improve the learning mechanism. When it comes to project evaluation, bilaterals may be different from multilaterals: UN agencies and GIZ have external funding agencies which usually commission the independent evaluation, while the bilateral such as USAID may conduct the evaluation by the agency since they contract out implementation to contractor companies.

## 2.2: Evaluation Types

Thus, the program design shall start with a needs assessment to draw the theory of change of the program. This will be followed by a baseline study that assesses the status of beneficiaries before taking the intervention to compare it with post-assessment evaluation which is the final evaluation. Practically, monitoring activities are also called midterm evaluation, where there are ongoing monitoring cycles to ensure the project is able to reach the output listed in the project's logical framework. Without a progress monitoring system, there is no way to get continuous feedback on each step and activity in the project (Thomas et al., 2021). Impact evaluation is assessing both the pre and post intervention phases. There is also the process of measuring within the intervention group and outside the intervention group which is also called comparison group to prove causality. This long duration enables impact evaluation to understand whether the program was able to achieve desired results and if it was a cause-and-effect relationship between

the project and participants or whether there were any other unintended consequences on individuals who were part of the program (Baker, 2000; Orr, 2018).

## 2.3: Evaluation Determinants

The determinants drawn to ease the pathway to robust evaluation are nine main steps that were extracted from the research through the primary and secondary data that was collected. The challenges and case studies gave a clear and practical understanding of what exactly should be done for effective evaluations: a detailed evaluation plan that clearly articulates the types and the purpose of the intervention to choose the convenient evaluation. When choosing the criteria based on which the impact of the project will be measured, there must be a way to prove causality between the intervention and impact. Coming in third, the timing of conducting the evaluation must be clearly stated from beginning to end. This is important since it is very difficult to decide to assess needs when the intervention has already started as well and the same goes for final evaluations which must take place after the services are finished and not during the intervention. Using a scientific sampling technique, tested and triangulated tools are also an essential determinante to ensure the intervention fits the context and community and to be able to generalize the study if this was the purpose of the intervention. It is important to understand the different objective functions for each partner within the partnership because each of them has their own agenda and interests. In that sense, ensuring accountability within development partner organizations, increasing ownership of public partners towards the indicators and efforts of the intervention and verifying that the evaluation body is independent and not hired by any of the other three partners will ensure that the agenda of the intervention is balanced between the

political and the scientific angles. Within the evaluation plan, there must be an agreed upon participatory approach that includes all the stakeholders within the correct time frame in the intervention that will help increase data accessibility and finally increase evaluation utilization and dissemination. This is the clear pathway framework for robust evaluation in development partnerships (Kabonga, 2018).

## 2.4: Purpose level

There are few different levels to consider when planning for an evaluation, it starts with asking why we are conducting this evaluation. The choice of assessment method depends on the intended use, with specific goals driving the procedure. The main objective at the policy level is to provide legislators with legislation that is supported by evidence. This justifies the adoption of an impact evaluation and calls for the use of a sampling technique capable of generalization and causal inference through experimental methodologies. On the other hand, the selection of the remaining three types of purposes is contingent upon the particular evaluation questions that require attention as well as the **OECD-DAC criteria** that have been chosen (OECD, 1991). In order to ensure a thorough and focused assessment that is in line with the established objectives, the type of evaluation becomes essential (OECD, 2011).

Figure 2 illustrates the visualization of the conceptual framework, the different purpose level, types of evaluations, criteria and expected outcome from the process without including all determinantes. It is simply a brief and concise description of them to better understand the complex cycle of evaluation.

Figure 2: Conceptual Framework: Pathway to Robust Evaluation in Egypt's Bilateral Projects



**Pathway to Robust Evaluation in Egypt's Bilateral and Multilateral Projects**
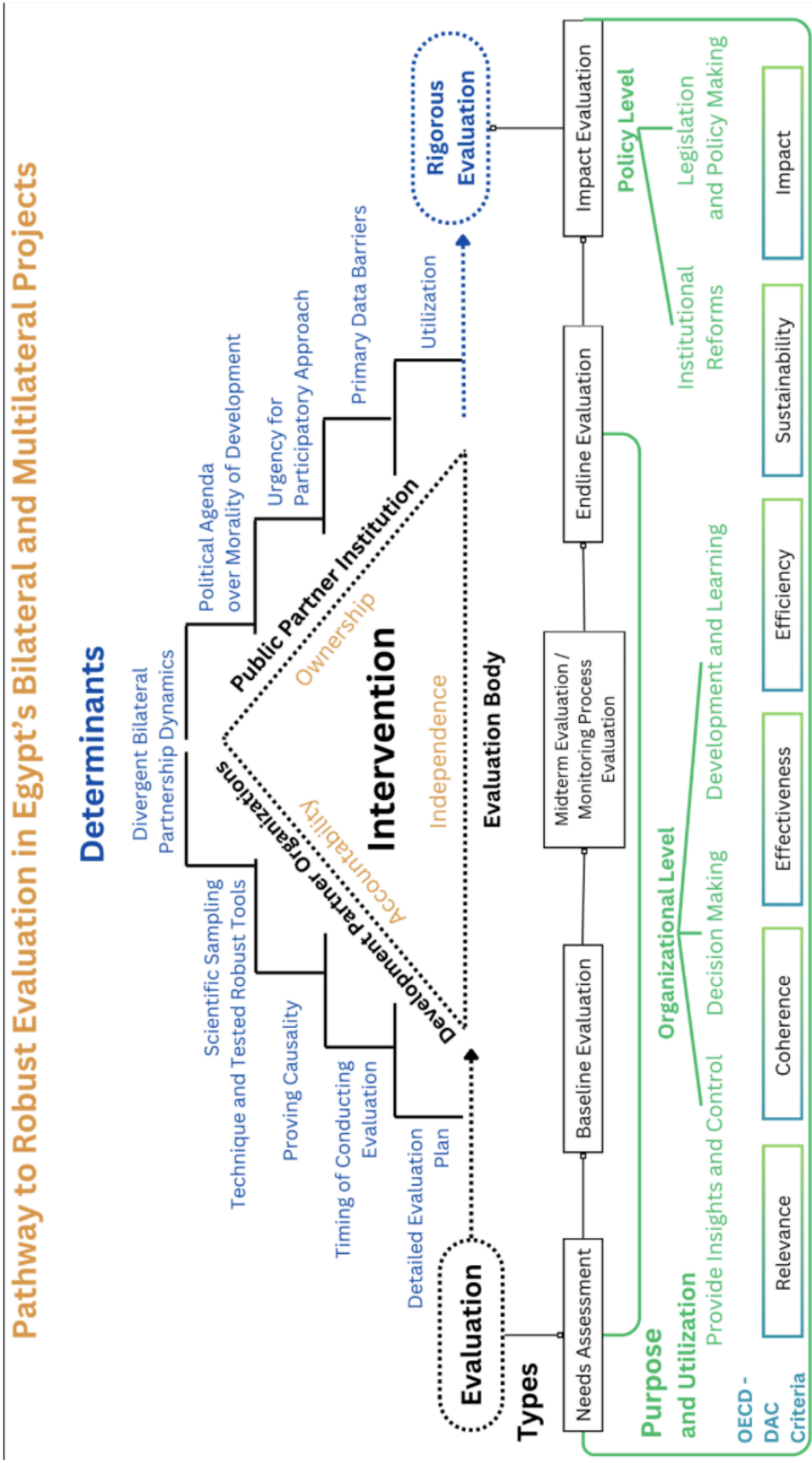
Sources: The Framework is inspired by (El Baradei et al., 2014; Ravallion, 2009; OECD, 1991; Orr, 2018; Field work findings Interviewee Eight, Nine and Sixteen, 2023)

# *Chapter Three: Literature Review*

The literature review is divided into three sections: section one **epitomizes evaluation for development projects beginning**, discusses **the concept of evaluation** in international development context as well as understanding how it looks in practice. The next section reviews the driving forces for international development and in specific bilateral and multilateral partnerships in an attempt to understand the context and challenges globally while also understanding the universal goals and the approaches for evaluation frameworks. The final section in the literature review is understanding evaluation in the **Egyptian context with a focus on three main development partners** working in Egypt in bilateral and multilateral partnerships with the government and exemplifying their current evaluation frameworks and practices used based on the Organisation for Economic Co-operation and Development (OECD).

## 3.1: Evaluation: In History and Practice

### 3.1.1: Evaluation Historical Framework

House (1994) theorized the definition of evaluation by understanding the underlying logic for evaluation: if there is x and y, then it is likely that one of them is better than the other; for example, y is better than x. Ranking them needs to be justified to the stakeholders who will benefit from the process and it is necessary to use the available resources to make the judgment. Orr agreed with House on the same logic and explained that the evaluation's primary responsibility towards a project is to make evidence accessible and foreseen and translate that evidence into words (Orr, 2018).

Ralph Tyler was the father of evaluation; he put *evaluation* into words by stating that it is the extent to which educational objectives have been achieved (Tyler, 1975). The literature in history of evaluation reveals that evaluation originated from education as the concept started as an integral part of the education process. That is why Scriven, in 1981, defined *evaluation* as "judging the worth or merit of something or the product of the process" (p.5). The exact definition is now used broadly by more researchers and evaluation institutions who agreed with Scriven's concept including the American Evaluation Association (American Evaluation Association, 2001; Kanyamuna & Phiri, 2019; Toscano, 2013). Researchers have contributed to the evolution of the concept over the years and contextualized it more towards the social program so that the definition of evaluation became the following:

> *Systematically and objectively assessing a policy, programme or project which is on-going or completed, its planning, design, execution and results. The main focus of any evaluation should be determining an intervention's relevance, efficiency, effectiveness, impact and sustainability with a view to appreciate its value, significance or indeed its worth (Kanyamuna & Phiri, 2019, p. 3).*

In terms of viewing evaluation within the framework of social programs, Edward Suchman was the first evaluator trying to unpack "black box theory" in relation to social programs (Astbury & Leeuw, 2010; House, 1994; Scriven, 1996). The theory of black box is concerned with evaluating the outcomes and effects of a program without understanding the logic of the program and its relation to the main program goal. It considers the program as a closed box where activities and inputs go inside and outputs come out from the other side while overlooking the risks, assumptions and logic that affect the program. Edward expanded on this theory by offering an antonym called the "clear or white box evaluation" which takes into

account program inputs, resources and expected assumptions that will yield positive outcomes at the end which makes evaluation more realistic and relevant to social programs' nature (Astbury & Leeuw, 2010).

Scriven has also stated that "evaluation is a very young discipline - although it is a very old practice" (Scriven, 1981, p.1). Evaluation as a practice started in the 18th century and evolved within sciences like education, management, and economics (Hogan, 2007; Scriven, 1996; Madaus et al., 2005; Toscano, 2013). Program evaluation in specific has undergone evolution through seven stages. Researchers have identified these stages as "reformation, efficiency, innocence, development, professionalization, and expansion/integration" (Hogan, 2007, p.3).

The starting point for program evaluation with the "reformation" as first stage (Hogan, 2007, p.3) began in 1792 when William Farish sought an objective examination method to assess his students' performance. He introduced the use of quantitative marking tools to aggregate scores and rank examinees, marking an innovation in the educational field. This idea paved the way for the development of quantitative questions that measure factual technical competencies as well as psychometrics, as opposed to qualitative styles (Toscano, 2013). Farish's pioneering work not only influenced education but also marked the first documented use of program evaluation as a standalone discipline (Hogan, 2007). The reformation of education by changing the method of evaluating students was just the beginning. The Powis Commission in Great Britain suggested that teachers' salaries should be determined based on students' performance in classes, indicating a growing focus on the evaluation of educational programs (Toscano, 2013). Formal evaluation

began to take root in two different areas: the military and education. In 1815, the Army developed a standardized system of policies to unify inspection, production techniques, and specifications, demonstrating the importance of systematic evaluation in diverse fields. The education sector saw further reform in 1845 in Boston, Massachusetts, as a system was established where schools and educational institutions were evaluated according to students' test scores to measure the effectiveness of the program offered (Hogan, 2007). Joseph Rice, an educational reformer, made an even greater contribution to program evaluation between 1887 and 1898 in his article entitled "The Futility of the Spelling Grind" (Colwell, 1998, p.25). He conducted the first comparative analysis to study the different ways of teaching spelling across school districts, producing findings that were announced in a forum in 1897. Rice's evaluation was the first formal educational program evaluation, marking a significant milestone in the evolution of program evaluation (Hogan, 2007).

Between 1900-1930, Frederick W. Taylor, who is called "the father of scientific management", dedicated his work about time and motion studies to explore horizons for production improvement and collecting data then evaluating current processes and test suggested alternatives (Locke, 1982). Maybe Taylor's intuition was not about M&E but his contributions to understanding efficiency made him a precursor to M&E approaches and evaluations in specific and to establishing the second stage in the cornerstone (Hogan, 2007). The third step focuses on Ralph Tyler, who is considered the father of educational evaluation. Tyler expanded on the work of Frederick and Rice by stressing the need to compare outcomes of 30 schools to evaluate program effectiveness (Hogan, 2007). He conducted the Eight-Year Study and concluded that internal comparisons between outcomes and objectives are more practical than comparisons

between experimental and control groups to avoid high costs and disruption (Madaus et al., 2005).

After World War II, the significant increase in developmental aid to build societies and provide facilities resulted in the neglect of accountability for the funds spent on those interventions, leading to what is now known as the "Age of Innocence." (Hogan, 2007, p.4) It was during this time that the formal development of M&E as a standalone field began (Hogan, 2007; Kabonga, 2018; Madaus et al., 2005; Toscano, 2013). Tyler's students further advanced his work by studying the relationship between outcomes and objectives, and by classifying objectives based on their relationship to each other and testing each outcome to ensure that their position in the hierarchy is correct (Hogan, 2007).

In 1965, "The Age of Development" (Hogan, 2007, p.5), Senator Robert Kennedy passed the "Elementary and Secondary Education Act (ESEA)" to ensure that governmental money was going to support underprivileged students and offer new development opportunities to help them. This was the first time public organizations admitted that evaluating programs can help direct money to the most critical areas, highlighting the importance of accountability and effectiveness in the use of public funds (Madaus et al., 2005). It was only the public who recognized the importance of practicing evaluation but also the donor agencies who decided to integrate the systems into their countries and UN level systems as a whole. UN agencies did not stop at that but also built their capacity to strengthen ownership towards these new systems and staff (UN Task Force, 1984).

Evaluation as a profession began to take shape in the 1970s, and a number of journals, institutions, and academic schools created their own specialized departments or programs with the primary goal of advancing the field and enhancing understanding, while promoting awareness and offering education to interested students (Hogan, 2007). The Evaluation Research Society established its own quarterly journal, in addition to other existing journals such as Educational Evaluation and Policy Analysis, CEDR Quarterly, Evaluation Review, and American Journal of Evaluation. Universities also recognized the importance of evaluation by offering courses in evaluation methodology. These universities include the University of Illinois, Stanford University, Boston College, UCLA, University of Minnesota, and Western Michigan University (Hogan, 2007; Madaus et al., 2005).

In the early 1980s, "The Age of Integration and Expansion" (Hogan, 2007, p.6) the field faced difficulties due to budget cuts under the Reagan administration. However, evaluation rebounded in the early 1990s as the economy improved, leading to an expansion and increased integration of the field (Hogan, 2007). Professional organizations and evaluation standards were established by a variety of associations who are still working until now in evolving the field of evaluation (Madaus et al., 2005). The evaluation efforts in developing countries focused mainly on tracking the delivery of inputs and activities rather than assessing the overall impact on beneficiaries. While many OECD countries were making progress in developing M&E approaches, there was still a lack of widespread adoption (UN Task Force, 1984). It is notable that the origins of evaluation as a field throughout history mainly traces back to the United States of America; many professionalizations, associations, authors, and institutionalization of the field originate from there (Kabonga, 2018). In addition to the American Evaluation Association

(AEA) that was established in the 1980's with over 3000 members now, the oldest and most ancient association (American Evaluation Association, 2001). Ten years later, countries started to gather their network of evaluators on the European side of the world and established the European Evaluation Society (EES). Then in Africa, the dominant evaluation association was established in 1990 and promoted country-level associations to start to work in their own countries for better engagement and capacity building (Kabonga, 2018).

Beginning in the twentieth century when educational psychologists started using treatment and comparison groups in their research methods to assess the different factors that affect their classroom circumstances, researchers found out that 14 to 18 percent of the studies conducted between 1914 to 1916 used comparison groups in their methodology (Orr, 2018). It took researchers a huge amount of time to be able to avoid bias while using these methods and to be ethically correct with treatment and comparison groups and take into consideration the time consumption and cost while choosing to conduct research with randomized controlled trials. In the late 1960s and 1970s, the first four research studies were conducted in the social world for programs related to cash transfers to low-income families which attracted funding agencies and the whole research community after witnessing the great results of the study that took a larger sample than any other experiment before which was five thousand families (Orr, 2018). The question highlighted from the experiment was if cash transfer programs would cause poor families to stop working or in other words have a negative impact on them instead of uplifting their lives (Orr, 2018).

Other researchers argued that the field started accidentally after the end of World War II, when developmental aid started being directed from governments of developed countries to developing countries. Governments started to ask about the impact of their funds and wonder whether their funds were really helping those countries become better or not. For twenty years, a network of evaluators started to evolve until the American Evaluation Association (AEA) was established in 1986 and then ten years later other countries started to gather their network of evaluators and established the European Evaluation Society (EES) (American Evaluation Association, 2001).

There are four main **reasons behind conducting evaluations** since the system started: to ensure effectiveness and efficiency for resources and objectives of the project; to provide a learning environment to understand the pitfalls of the past and ensure better informed decision making in the future; to engage stakeholders in a transparent mode that will enable trust and support; and to assess if the interventions are able to make a meaningful difference in people's lives. The complexity of the issues facing the world in 2023 whether they be environmental, social or economic leads to the designing of programs that are also complex in their objectives which makes the role of evaluation all the more important (EvalCommunity, 2023).

### 3.1.2: Debates on Good Quality Evaluation

In practice, evaluation is divided into three main aspects: time of evaluation, who conducts the evaluation, and the selected methodology used for evaluation. The common ground between evaluation and social sciences was noticeable in the literature; however, the differences are undeniable. Scriven, whose publications in evaluation studies drew the first notions in the

field and stapled the origin concepts, stated that evaluation "determines the merit, worth, or value of things" (1991, p.5). On the other hand, social science research depends on empirical evidence to find factual results that are considered the conclusion of the study. However, in evaluations, the values or standards are the engine that runs the empirical techniques from social science research to find meaningful conclusions that integrate with the standards that were in the evaluation equation (Scriven, 1981).

The characteristics of evaluation and social science research put it more within the realm of a profession rather than a discipline. As evident in the literature on the history and evolution of evaluation, social science research considered evaluation to depend heavily on values, which is neither scientific nor objective. As the professionalization yielded better methodologies and results over the years, the discussion is still ongoing to appraise it as a standalone discipline (Coffman, 2004).

Evaluations can be divided into formative and summative categories based on temporal factors. Formative assessments are essential when a new project or phase starts because they offer important insights during the early stages. Summative assessments, on the other hand, are used to determine whether to continue or end a program. These evaluations explore the project's short-, intermediate-, and long-term results. Summative evaluations thoroughly analyze the project's overall outcomes, whereas formative evaluations mainly aid in understanding the project's output level.

For many researchers, evaluation is a broad term that encompasses multiple processes, and all of these systems together can be referred to as evaluation. Baker's theoretical framework

for comprehensive evaluation incorporates "**monitoring, process evaluation, cost-benefit evaluation, and impact evaluation**" (Baker, 2000, p.10). **Monitoring** is a vital component within the evaluation process; it is a continuing feedback loop through program implementation that enables identifying problems and gathering data for future evaluation to measure if the project was executed as designed (Baker, 2000). As for **process evaluation**, it is quite similar to monitoring. However, it is more concerned with 'how' the program works and is executed. Kanyamuna & Phiri in 2019 clarified the determinants of evaluation, two of which are effectiveness and efficiency. Earlier in the 2000s, Baker identified **cost-benefit evaluations** as a standalone type that compares project resources to their alternatives usage to estimate whether a program was able to use resources to their maximum capacity to produce maximum benefit for targeted beneficiaries. Outcome harvesting is observing change and pinpointing what you have done or what was your contribution to this intervention: you observe change and then go back to see how you contributed.

**Impact evaluation** is more focused on people, it determines whether the program affected individuals and households in relation to program intervention attributes. While measuring consequences, impact evaluation tends to explore whether they are unintended or intended and whether they have negative or positive impact on targeted people (Baker, 2000). A number of essential elements are involved in programme evaluation, each with a specific function. Usually, the process starts with a needs assessment to determine and comprehend the target population's requirements (Koolwal et al., 2010) An essential element of programme evaluation is impact assessment, which seeks to ascertain the program's efficacy by providing an answer to the following question: "what would have happened in the absence of the

31

programme?" (Baker, 2000, p.155) This entails comparing the program's results to what could have happened in the absence of it. It is also important to establish a counterfactual, which is typically accomplished by assigning people at random or by choosing a group that is unaffected by the programme (Baker, 2000; Scriven, 1996). Could the programme or project be improved in order to achieve the desired results? Are resources being used efficiently? These are the kinds of questions that can only be answered by undergoing an impact evaluation, which is a method of measuring the outcomes of a programme intervention in separation from other factors involved (Baker, 2000). There is a growing trend toward evidence-based approaches in policy making and project funding. This necessitates a heightened emphasis on impact evaluation to provide substantiated data that can inspire informed decision-making. However, 71% of non-profit funders provided no support for their organization programs to be evaluated rigorously although impact evaluations fitted perfectly in this specific case according to a study by the Center for Effective Philanthropy (Roob, 2014).

## 3.2: International Development Goals Framework

It is essential to take a holistic look at international development in order to fully appreciate the complexities of evaluation. In this context, evaluation is deeply entwined with a variety of interconnected factors, including cultural subtleties, socioeconomic dynamics, and geopolitical considerations. Examining international development offers an essential context for understanding how assessment framework's function, evolve, and support the main objective of promoting constructive change globally. Through navigating the terrain of international development, one can acquire insights into the opportunities, challenges, and changing

paradigms that influence the evaluation processes that are essential for sustainable progress and well-informed decision-making.

Every year billions of dollars are spent on official development assistance (ODA) from donor agencies in developing countries. Yet, little information is known about the real impact of these interventions and whether they met the actual needs of beneficiaries (Baker, 2000; Koolwal et al., 2010; Roob, 2014). One of the key historical points in international development history is the UN Millennium Summit in 2000 which included 189 UN state members and led to the development of the Millennium Development Goals (MDGs). MDGs are eight goals tackling poverty and hunger reduction, education inclusivity, gender parity, elimination of child mortality and increasing women empowerment in poor and developing countries (UN, 2000). The Sustainable Development Goals (SDGs), as a successor for MDGs, focused on the whole world, whether rich or poor countries, to address 17 goals from 2015 to be achieved by the year 2030 (Dang & Serajuddin, 2020; McCloskey, 2015; UN, 2015; Yonehara et al., 2017). One of the main drawbacks of MDGs was the lack of evaluation methods for the targets and their indicators. To remedy this, Article 71 of the SDGs exclusively mentioned goals centered on data and evaluation: "they will be rigorous and based on evidence, informed by country-led evaluations and data which is high-quality, accessible, timely, reliable…" (UN, 2015, p.31). While the main objectives of the SDGs may seem to be quite similar to those of the MDGs, they are different as they are more inclusive and extensive with 169 targets versus 18 and 232 indicators versus 60 as they tackle additional problems related to the environment, economic growth and social inclusion (McCloskey, 2015).

Vandemoortele critiqued SDG targets by saying they lacked one or more essential elements that should be part of any written target, that they should contain outcomes in numbers, a deadline, and structured domain. For example, in item 16.5 which aims to "substantially reduce corruption and bribery in all their forms" (UN, 2015, p.21) the target lacks the three aforementioned elements. Vandemoortele claims that only 30 targets were formatted correctly, while the remaining are fluid (2020).

However, other researchers have reached the conclusion that the SDGs are deliberately broader in scope than MDGs as they are left for each country to customize according to country-level needs, targets, and indicators. Additionally, countries are required to report back to the UN in intervals (Dang & Serajuddin, 2020; McCloskey, 2015; UN, 2015; Yonehara et al., 2017). In a sense, this enables international development and aid to be more effective since programs will be based on goals that the country has determined are a priority to work on and the indicators are reported back to the development agency, government level and UN committee as part of the SDGs evaluation framework agreed on a national and global level. The UN Evaluation Group in 2016 agreed on an evaluation framework to be implemented on a national level to ensure that gathering the data on a global level is refined. The evaluation framework depends on three major phases, each phase takes 5 years to complete. The first phase is concerned with ensuring that international organizations and governments are planning their programs according to the SDGs and that they are contextualized according to each country's needs and that funding is available for these huge projects. In this phase, evaluability assessment and proactive evaluation take the lead to ensure needs gaps are filled and to prepare for program implementation. The following phase is about implementation, modification and enhancement in

which projects use monitoring approaches and formative evaluations. The final phase where the SDGs time frame comes to an end, projects are nearly finished, and essential outcome and impact evaluations take place to ensure on the national and international level that the data needed to serve as an engine for the future agenda is available and ready (Dang & Serajuddin, 2020).

One of the earliest aid programs was the "Marshall Plan" after World War II. Between 1948 and 1951, the United States dedicated 12.5 billion dollars to support Western Europe in rebuilding their countries. The aid program proved successful as it contributed to European countries' development and growth, enabling them to purchase the basic supplies needed to re-construct (Crafts, 2011). According to the Institute of Development Studies[1], an international aid and development have passed through three phases: the period following World War II, mainly the post-colonial world of the 1960s; the end of the Cold War; and the global economic crisis that began to take shape in 2005 (Eyben, 2013).

For Eyben, the first phase, which was between the 1960s to 1990s, was the result of political and economic difficulties that were faced by emergent nations because of the post-colonial world. In the period of the Cold War, the difference between countries within the global order became increasingly clearer. There were developing countries that were in continuous internal and external conflicts that challenged their military and economic situation and there were powerful developed countries that could provide development assistance to strategic allies (Eyben, 2013). Phase three, which began in 2005 and continues until the present

---

[1] Research Center founded 1966 affiliated with the University of Sussex, in Brighton, England

time, is defined by the global economic crisis that turned the world into a "multipolar" between different powers of countries and traditional donors (Eyben, 2013, p.2).

The order laid out by Eyben on the evolution of international development and aid complements the findings of Dichter which points out the establishment of the "Point Four" program proposed by Harry Tuman in 1949 which aimed to provide technical assistance to emergent nations. This program along with the inaugural speech and law formation that accompanied it marked the beginning of the Technical Cooperation Administration, which was later changed to USAID (Dichter, 2002). The Marshall Plan reinforced the idea of how external aid can support countries to grow to be successful; it aimed to support Europe to return to its previous glory. It provided essential commodities, rebuilding infrastructure, and improvements directed towards restoring health and education to their pre-war conditions. Much in the same way, development assistance that came in the years after that until the present day has strived to accomplish the same aims albeit in a more complex way. It continues to try to address the world's problems by alleviating poverty, improving medical infrastructure, increasing education levels, addressing environmental dilemmas and other issues that have evolved in recent years (Dichter, 2002).

## 3.3: International Development Debates and Challenges

Development is a complex and deeply interconnected field as is illustrated in the following quote from Waldrop:

> Except for the very simplest physical systems, virtually everything and everybody in the world is caught up in a vast, nonlinear web of incentives and constraints and connections. The slightest change in one place causes tremors everywhere else, We cannot help but disturb the universe (1992, p.64).

Current world problems are **complex** and interconnected between the societal and economic systems. Attempting to solve one aspect in one system can have unpredictable consequences on other fragments of the same system or even on other systems; program managers may not be able to foresee the impact of their intervention, particularly if there's a negative side effect, and may have only focused their efforts on improving the quality of life for people in developing countries. Hence, the need for thorough planning, monitoring and evaluation has become mandatory to prevent the causing of any harm as well as to measure planned outcomes and impact of programs (Dichter, 2002; Ravallion, 2009).

One of the major drawbacks of receiving foreign assistance in the name of development is **increasing dependency** in those emerging countries (Dichter, 2002). Eyben agrees with Dichter, stating that international development is about politics and power battles between countries and that it gives more power to developed countries over developing ones (2002; 2013). The dynamic began with emergent nations needing money to overcome a specific obstacle such as starting over after war but the need was never fully satisfied - perhaps due to misallocation of funds or ineffective interventions - and they continued to be in need of support, assistance and money from developed countries.

The World Bank has emphasized this notion by conducting research focused on countries who received the greatest amount of aid and came away with the conclusion that they often performed much worse in comparison to their counterparts. The study found that aid has increased dependency as well as reinforced policies to keep receiving funds and, in some cases, intensified **corruption** (Lele, 1990). Duflo, E., & Banerjee, A. V. have also echoed this by

presenting a case in which aid inadvertently increased corruption in governments and reduced their ability to solve their country's problems (2011).

Dichter argues that **development has become a business** that uses marketing tools and objectives more than a way to help developing countries. Practitioners involved in development often professionalize their intervention and present it in a branded way to donors, even though it may not be genuinely helpful to the intended beneficiaries (2002). This false dichotomy that was speculated by Dichter has since been disputed by several researchers as one of the main purposes for structuring **monitoring and evaluation** systems for development program is to **enhance accountability and transparency** along with measuring the real impact of these programs compared to their planned objectives (El Baradei et al., 2014; Ravallion, 2009; Marshall & Suárez, 2014).

However, Ravallion perceived the dilemma in a different way and called it a **"myopia bias"** (Ravallion, 2009, p.32), a phenomenon in which practitioners tend to prefer short-term projects that yield information on short impact interventions over accepting the fact that fixing world issues takes time and poverty cannot be alleviated "right now". The urgency to find solutions and evidence from real programs while simultaneously favoring quick results is not realistic nor productive and this is exactly what practitioners need to understand (Dichter, 2002; Ravallion, 2009). Funders usually encourage development partner agencies to stay visible and report immediate results which makes these agencies inclined to pursue shorter interventions to align with funder expectations instead of long-term objectives that may or may not be achieved at the end (Holzapfel, 2016).

When it comes to development stakeholders, there are many actors at play and mapping it out can be quite complex. Therefore, it was an essential step for development organizations to attempt to identify the different types of stakeholders and map out their accountability towards them: upward accountability to funders and downward accountability to clients (Holzapfel, 2016; Liket et al., 2014).



Figure 2: Illustrated by author from Holzapfel 2016; Liket et al., 2014: Accountability framework for Development Stakeholders

The fear of not achieving intervention results impacts the whole development system: upward accountability to funders creates risk aversion and lack of creativity in new interventions. In an example by USAID, Natsios showed that funders have the power to stop sponsoring a project if it failed, or if they received a negative evaluation report. Hence, staff members are usually reluctant to innovate or brainstorm new programs (Holzapfel, 2016; Natsios, 2010). Due to the interest in getting usage of each dollar spent, funders are likely motivated by getting the most out of public goods to increase their impact in society. It may even be the case that an evaluator's own work may be used as proof against them to stop projects if the results are statistically significant that the program proved not to empower women to the extent that it initially set out to for instance (Liket et al., 2014). Between the aid optimists and pessimists, aid had little direction as to what could be the solution. The same question has been asked for

decades whether aid works or not ; But instead of asking the question in this format, it can be reframed to: "*Which programs work best, why and when? – Which concepts work, why and when? How can we scale up what works?*" (Sachs, 2006, p.226).

## 3.4: OECD - DAC: Between Evaluation and Development

The OECD - DAC - Aid Effectiveness report from 2003 emphasized the critical role that evaluation plays in raising the caliber and efficacy of development cooperation. Experts in bilateral and multilateral evaluation were brought together by the DAC Working Party in Aid Evaluation, a special international forum, to improve procedures and learn from collaborative projects. The Paris Declaration of 2005, which represented the first agreement between donors and recipients on the fundamental values of ownership, alignment, harmonization, results, and mutual accountability, kept the momentum going. The 2008 Accra Agenda for Action, which emphasized donor support alignment, streamlined in-country efforts, shared accountability for attaining specific, measurable goals, and aid recipients' autonomy in creating national development strategies, was widely embraced. Another common tool for evaluating performance, Results Frameworks, was highlighted in the ensuing 2011 Busan Partnership. These frameworks are based on a manageable number of indicators that are in line with the priorities of developing nations. These ideas came about as a result of a group effort to comprehend the obstacles that aid-related development is facing and to step up efforts to reach the lofty goals established by the Millennium Development Goals (MDGs) (OECD, 2019). The Aid-effectiveness working group was the main author for the six-main principles of evaluation which are effective, efficient, impactful, sustainable, relevant and coherent. The principles are working mainly on aid effectiveness forums to achieve better development results.

## *Chapter Four: Methodology of Research*

In chapter three, I examine the research method selection and data collection techniques. Then, the sampling strategy was identified. Following that, the study framework shed light on the issues that arose in the field and the role of the researcher. Finally, a summary of this study's limitations and ethical issues is provided.

### 4.1: Methodology Selected and Design Rationale

In order to assess evaluations in Egypt, the status quo, opportunities, challenges and ways for better methods to be used, it was hard to find enough data about evaluations since this field is still under-researched. Thus, tapping into this research needed a method that can support exploring the on-ground practices deeply. That is why a need for qualitative data was needed to identify the broad spectrum of variables and find relationships between them (Lune & Berg, 2016). Key informants from donor organizations, independent evaluators, researchers and government officials were targeted to conduct interviews since they are the main stakeholders of the study. Given the macro-level lens and research purpose, there was no need to reach out to intervention beneficiaries. Given the design rationale, the questions were inspired from the literature review and research background and were given an exploratory nature of in-depth and open-ended questions to give the respondent the opportunity to provide case studies and examples from their experiences. The questions tackled the evaluation purpose, plan, cycle, stakeholders, challenges in conducting and their perception of rigorous evaluation determinants while giving examples from their work sites (Abdelhamid, 2005; Baker, 2000; El Baradei et al.,

2014; Hogan, 2007; Scriven, 1981). In that sense, the respondents are all anonymous and confidential (Lune & Berg, 2016).

## 4.2: Data Collection Process

The collected data depended mainly on in-depth, semi-structured interviews with donor staff affiliated with the evaluation process, independent evaluators and researchers of developmental economics as well as government officials. To ensure triangulation and verification of data gathered, I used content analysis of official documents shared by respondents that were not available online to increase validity and reliability. In addition to that, I used the observational method by attending a few evaluation and learning events that took place in Cairo in 2023 by German and American agencies to witness what is currently happening in the field. I was also able to observe, through my work as an external consultant on a project for an international NGO that aimed to develop a few evaluation tools, the types of challenges that my colleagues and I would face during implementation and the true responses of stakeholders (Lune & Berg, 2016). The depth of understanding that can be obtained from an investigation is increased when different study design techniques and theories are employed (Marshall & Rossman, 2014).

The process of interviewing such elite stakeholders is a very efficient method to use since most of the selected participants have more than 10 years of experience in evaluation and international cooperation projects. Because those individuals were able to offer me knowledge and confidential information about project design, explain the rationale behind the selection of evaluations, and recount the mistakes that continue to take place in the field until the present day,

their support was invaluable as not all information about the Egyptian context can be found through secondary sources.

The data collection process took me two months to complete fully. The interview duration was between one to two hours. All interviews were scheduled for 45 minutes only, but given the interviewees eagerness to share their knowledge with a research project, the majority shared their time generously in order to offer more examples and documents. All interviews except one were recorded on my personal laptop and stored on it. In adherence with IRB conditions, this was done after their hand-written or email approval for the consent was shared with the IRB committee. All the interviews were one-on-one interviews and all of them were online given their busy schedules and because four of them were outside Egypt at the time of interviews.

Interviewees included four government officials working in bilateral and multilateral projects, ten staff members with different seniority levels in M&E department, two working for UN-agencies, three from the German agency and three from the American agency. In addition to that, a total of ten independent evaluators working closely with the same projects were interviewed. To ensure diversification, I also interviewed four researchers working for research centers and involved in evaluating developmental projects in Egypt. It is important to mention that the total number of interviewees were nineteen. Since the tool was intended to ask them about their last 5 years experiences, I was able to talk to them through different roles they were in their last few positions.

Table 1: Interviewees Background

| Interviewees with background in | Government Officials | M&E Staff Members | Independent Evaluators | Researchers | Total Number |
|---|---|---|---|---|---|

| Overall without repeating | | | | | 19 |
|---|---|---|---|---|---|
| USAID Projects | 1 | 3 | 3 | - | 5 |
| GIZ Projects | 1 | 3 | 3 | - | 5 |
| UN Projects | - | 2 | 4 | - | 4 |
| OECD Experience | 1 | 2 | - | 2 | 2 |
| Research Centers | 1 | - | - | 2 | 3 |
| Repeated | 4 | 10 | 10 | 5 | |

*Note: The table indicates the overall experiences but not their numbers as in the last five-years it occurred that they have worked for several organizations/ experiences.*

Table 2: Interviewees Years of Experience

| Years of Experience | Around 15 Years | Around 10 Years | Five or more |
|---|---|---|---|
| Number of Interviewees | 4 | 8 | 7 |

To ensure using robust tools, I piloted the interview questions with three participants before starting the actual interviews to ensure validity and reliability of the questionnaire and that the questions are easy to understand (Marshall & Rossman, 2014). This helped me review the tool in different rounds and with interviewees of different backgrounds to ensure that the tool can respond to the wide spectrum of the evaluation scene in Egypts' bilateral and multilateral partnerships.

The interviews followed Marshall and Rossman's (2014) guidelines for guided interviews with semi-structured questions. In this type of interview, the participant is given the opportunity to steer part of the conversation and shape answers in the direction they choose by the researcher as they delve into a variety of topics to get the participant's viewpoint (Marshall & Rossman, 2014). The style of in-depth interviews made it easier to address the most important issues raised by the study. I chose not to allow the conceptual framework to direct the interview questions, even though I had done extensive literature reviews prior to the interviews. After the interviews

were finished, the conceptual framework was chosen, but not before a thorough comprehension of assessments in bilateral partnerships was established. In order to reduce researcher influence and reveal participants' points of view, more open-ended questions that were less attached to conceptual frameworks were used (Marshall & Rossman, 2014).

## 4.3: Sampling Identification

The sampling methodology employed in this study is purposeful and guided by theoretical considerations. The snowballing technique is a non-probability technique that was utilized to reach out to diverse study participants, aligning with the insights gathered from prior literature review and the research question. Therefore, employing the snowballing method was the most effective approach to identify individuals with specific attributes or characteristics required for the study (Lune & Berg, 2016). I started to gather contact information and acceptance to participate in my research while attending the evaluation and learning events for USAID and GIZ and was invited to them throughout my personal network from academia and professionally. I also attended online webinars held by the same agencies and a few others abroad to be updated with evaluation news that was registered through a variety of subscriptions to emails and newsletters related to international development and aid-effectiveness. I was able to meet professors in development studies in different universities in a few events throughout the year where they introduced me personally to a few other colleagues who are very experienced in the field of evaluation and have themselves conducted robust evaluations. I also acquainted myself with their papers and work that are related to evaluation in the MENA region and in specific Egypt with either J-PAL, IFPRI and other research institutions.

A number of the interviewees offered perspectives on noteworthy bodies and individuals, broadening my understanding beyond what I had previously learned. It is crucial to remember that sample representativeness is not the goal of qualitative research, including this thesis (Lune & Berg, 2016). Rather, important patterns are revealed by the variety of informants providing in-depth details about current circumstances and difficulties. Nineteen interviewees made up the sample size, which was chosen by taking into account both practical saturation and theoretical considerations. By this time, I had transcribed all of the available data, indicating the saturation point of information, and I had encountered situations where interviewees mentioned sources I had previously met.

## 4.4: Data Analysis

I used open coding to delve into the qualitative data and generated 97 codes that started to merge and find points of intersection between each other until reaching 22 codes that were themed up into four main themes and inside each one were three sub-themes. Using qualitative in this study to explore the research topic in depth and present new ideas. In contrast to traditional hypothesis-driven research, taking an inductive stance derives theories straight from the data. Up until theoretical saturation, the iterative method of analysis and theory development continued, suggesting that more data would not yield fresh insights. This approach relied heavily on the inductive coding technique known as "ground-up coding," which allowed the codes to develop naturally from the data without the need for preconceived assumptions. This method is especially helpful for exploratory research and for developing new theories that offer a complex and contextually rich knowledge of the topic (Marshall & Rossman, 2014).

I used Atlas.ti website to analyze the data after transcribing all interviews in google documents. I imported the transcription after removing all identifiers of participants and started analyzing each sentence in iterative cycles that took two months of analysis, ensuring an unbiased and transparent method. The main themes were chosen since they were the most repeated and can include a few codes and sub-themes that are the root cause or result of the central theme. As shown in Figure 4, these 24 codes were at least repeated ten times in the interviews and were significantly shown in the analysis to be used in the findings section. I then segregated the codes, sub-themes, and themes into more concise titles and ideas that can flow from each other and create a meaningful story for the analysis.

## 4.5: Limitations of the Study

The study's limitations stem from its exclusive focus on Egypt and its evaluation of development projects within the country. High-ranking staff members and independent contractors were the main subjects of the interviews, which offered a top-down analytical viewpoint without exploring the perspectives of the beneficiaries. Examining grassroots viewpoints may improve the breadth of subsequent studies. Also, interviewing staff members from headquarters might give the research a broader perspective that did not happen due to time limitations. The study's sample size is another significant limitation, although it was chosen based on data saturation considerations. Interviewing a wider variety of staff employees may have produced more insightful results.

## 4.6: Ethical considerations

Ensuring the validity and reliability of my thesis through observer and method triangulation involves reaching out to various stakeholders (Ambert et al., 1995; Neuman, 2013). Adhering to all Institutional Review Board (IRB) rules, I obtained approval before initiating the process and upheld participants' safety and rights. Before interviews, respondents were free to opt out at any point, and verbal and written informed consent was obtained (Miller & Cresswell, 2010). Ethical considerations were a priority, and the interview guide and study received approval from the American University in Cairo's IRB on April 27th, 2023, ensuring compliance with ethical standards. To maintain confidentiality, interview data was treated as confidential at all research stages, and anonymity was preserved in the final write-up and transcripts, using pseudonyms for respondents. Throughout the research, I transparently communicated the purely academic nature of the study to respondents, engaging with them solely as a master's degree candidate and avoiding conflicts of interest within my employment and personal professional networks.

# *Chapter Four: Policy Context*

In Egypt, when the revolution happened in 2011, there were greater reasons to support the rise of the field of evaluation. Government entities and donor organizations thrived to report achievements that meet people's expectations (El Baradei et al., 2014). Not only that, but it was also beneficial to know the status of the projects and where the huge funds went and to what extent they were impactful. The Ministry of International Cooperation was the first government organization to establish an M&E unit to gather information on the number of developmental projects working in Egypt, including the areas of work and the fund spent on each. As for impact evaluations, the Abdul Latif Jameel Poverty Action Lab (J-PAL) was launched in June 2003 with the aim of reducing poverty by providing science-based evidence research  (Dhaliwal, 2022). The Ministry of Planning and Economic Development (MPED) partnered with Community Jameel and J-PAL to launch the Egypt Impact Lab in 2022 that aims to work closely with the government, donors, researchers and development practitioners to provide rigorous impact evaluations in Egypt with a focus on poverty alleviation and social protection, employment, gender empowerment, and family development. Not only that but also there is the Egyptian Evaluation Association that is part of the MENA region Evaluation Association, both are working on ensuring that evaluators are backed up with a community working in the field (EvalMENA, 2024).

Aspects of development projects that are separate but related are monitoring, evaluation, learning, and accountability. Even though improving internal accountability and efficiency within development agencies is the main goal of aid project evaluations, governmental counterparts—who are vital players in guaranteeing the effectiveness of aid—own little of the

evaluations and don't fully comprehend them. Data scarcity and reluctance to publish evaluation results are partly caused by the opacity surrounding the political economy that supports aid and development initiatives. Monitoring, evaluation, learning, and accountability are all interrelated in the larger context of development initiatives, and this is why there is an increasing need for more structured and organized evaluation processes to address these problems (Valadez & Bamberger, 1994).

The three selected partners working in Egypt that have been chosen for the study are the U.S., Germany and selected UN agencies. Hence, taking into account the projects and themes they are working on was an essential part of reviewing the literature to understand the holistic approach of the international development scene in Egypt.
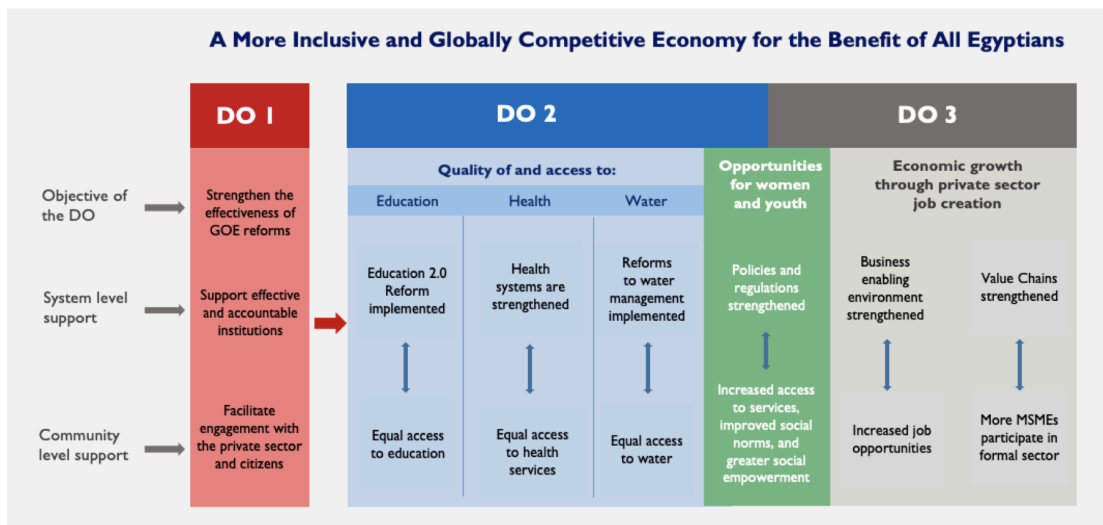


Figure 3: USAID Country Development Strategy 2020-2025

The U.S. works in Egypt through its agency USAID, which works with different implementing partners to achieve the 3 D.O.s in figure (2). The main intermediate results are to improve the health of targeted beneficiaries, improve education quality, provide better water and

sanitation services, increase women's social empowerment, and accelerate economic development. The agency works through various implementing partners in each sector, ensuring resources are conveniently and geographically correct (USAID Egypt, 2020). The USAID evaluation policy emphasizes evidence-based decision-making to enhance the impact and effectiveness of development programmes. It lays out a dedication to thorough assessments that support accountability and learning. The goal of the policy is to guarantee that assessments are methodically organized, carefully thought out, and usefully applied to direct subsequent actions and policies (*Evaluation Policy | Strategy and Policy | U.S. Agency for International Development*, 2020).

The United Nations has maintained a large presence in Egypt, with 32 agencies, funds, and projects currently operating there. With a focus on sustainable development, the UN offers beneficiaries direct assistance and institutional capacity building. The 2018–2022 United Nations Partnership Development Framework (UNPDF), which is in line with Egypt's Vision 2030 and the global SDGs, defines the partnership with the Egyptian government. With a $1.2 billion budget behind it, this partnership is a result of a deliberate attempt to make a major contribution to Egypt's development objectives (UN, 2018). The United Nations Development Programme (UNDP) uses the "UNDP Evaluation Policy" as a basic framework to direct and improve its evaluation procedures. This policy, which was released by the UNDP's Independent Evaluation Office (IEO), describes the guidelines, expectations, and practices that govern assessments aimed at enhancing the efficacy, efficiency, and significance of the organization's interventions. It highlights how assessments support accountability, organizational learning, and evidence-based decision-making. The policy integrates evaluation practices throughout the programme lifecycle,

encouraging a culture of continuous learning and adaptive management. All things considered, it is essential to upholding high performance standards, guaranteeing stakeholder accountability, and optimizing the effectiveness and long-term viability of UNDP's development initiatives (UN, 2018).

Since 1956, the German government, in collaboration with Egypt through the German Federal Ministry for Economic Cooperation and Development (BMZ), has spearheaded initiatives on behalf of multiple German ministries and the European Union. Executed by GIZ, these projects are in harmony with Egypt's Sustainable Development Strategy (SDS): Egypt Vision 2030, concentrating on crucial domains like climate, energy, environment, and natural resources. Currently, there are six active projects in these areas. Furthermore, efforts extend to training for Sustainable Growth to enhance opportunities in the job sector, with thirteen ongoing projects (GIZ, 2023). As per the literature, there is no open-source pathway to reach GIZ evaluation policy, however, interviewees illustrated the tools used in M&E and how they got improved over the years (2023).

Analyzing the available data shows that only USAID is sharing its most updated framework working right now and the main results they aim to achieve by 2025. While GIZ, no country profile shares the ongoing efforts in detail to understand and compare their plans to what has been implemented. Only one-page fact sheets for each project that do not combine in one agency profile nor understand their holistic approach aligned with Egypt's vision 2030. Regarding the UN, the only available country document is dated 2022, with no shared document for what is after that since we are approaching the end of 2023.

## *Chapter Six: Data Analysis and Key Findings*

The chapter analyzes the findings from the nineteen interviews conducted by the organizations selected for the study. The results are divided into four main themes: (1) Stakeholders Dynamics Challenge; (2) 'Impact does not happen overnight': Challenges in measuring projects' impact; (3) Purpose-related challenges and opportunities; and (4) Evaluation Utilization.

Through the first theme, the findings are curated around the implications of currently designed interventions that yield to the inability to measure the impact of projects since they are designed to be short, repetitive and redundant in some cases out of a set of ideologies and ignorance around evaluation and its nexus with bilateral and multilateral partnerships. The second theme is going back to broader evaluation types and comparing staff members and other stakeholders between going through a check-the-box or comprehensive evaluation and what other purposes could be essential to identify and choose the most appropriate evaluation type. Third are the stakeholders in those partnerships, the roles played by the funder, evaluator and implementer, and how the project agenda manifests the overall driving forces of evaluation. Finally, how to best utilize evaluations and the challenges encountered that make utilization of evaluation hard to happen when data is insufficient, stakeholders need to be made aware of evaluation significance and the learning opportunities that will come from evaluations.

The examples and status quo given and inspired by field work was from projects that are working in education reform, women empowerment, enhancing refugees' livelihoods, private sector engagement and social entrepreneurship. The organizations are mainly GIZ, USAID, WFP, UNICEF, UNHCR and UNDP. All interviewees had at least 5 years of experience in the

evaluation field and have worked through the projects and organizations for at least one-year to be eligible to provide their feedback. For few organizations and projects, using triangulation and interviewing two sides from the stakeholders was essential to understand the full-picture.

## 6.1: Stakeholders Dynamics Challenge

Having grasped through the primary challenges and opportunities in evaluations without attributing sole responsibility to any specific party, this section delves into the distinctive roles of each stakeholder. Exploring these diverse perspectives is critical for achieving enhanced evaluation results and overarching developmental objectives. The focus will be on the three key participants in bilateral and multilateral projects in Egypt, elucidating each partner's role and level of engagement. Additionally, this section will draw comparisons between evaluations and research, elaborating on how to extract the utmost value from each.

### 6.1.1: Conflicting Objectives of the Main Players

Bilateral and multilateral projects are well-known in Egypt to be between development partner organizations and public partner institutions that are part of the government. This happens within the framework of GIZ and USAID who partner with specific ministries as well as the UN whose agencies partner with the government in achieving different goals. The development partners normally choose the ministries or institutions that are affiliated with the area they aim to tackle either in health, education, environment and so on. Each side of
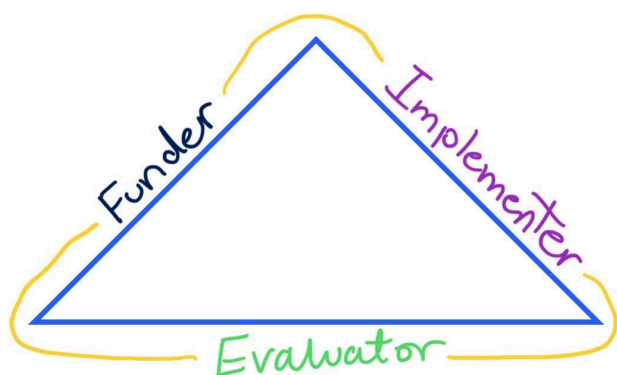
the triangle in the figure has its **own objective functions** and works collaboratively. However, based on the primary data that was collected, what tends to happen is that each side works only with their theory and interests and does not really take into account external factors and stakeholders. There are a lot of restrictions, challenges and limitations related to the public institutions body. This counterpart has its own structure, priorities and scope.

**Starting with the funder,** whether it be the German or American development partner organization or the UN or EU, all obtain their money from taxpayers in their countries as per the stakeholders dynamics framework (Holzapfel, 2016; Liket et al., 2014). This gives them the authority to choose the project objectives and scope. **Next comes the implementer** and, in our case, that is the public partner who takes the responsibility of creating the project implementation plan according to the project design by the funder organization. Interviewee Nine who illustrated the relationship between the partners as per the figure drawn stated the following:

> *"This is the point at which the constraints of each player come into play. The donor would say: 'We have a lot of money. We want you to work in ten governorates.' And so money is not a problem anymore. It's just the capacity of the implementer. The money is there, but they just don't have enough workers. They don't have enough networks. They don't have operational capacity. They don't have the logistics for it" (Interviewee Nine, Development Economist, May 2023).*

Through the analysis, it was discovered that it is not about the capacity of the implementer only but also about the inclusivity of all the partners altogether. At this point it becomes important to ask: **have we included the main stakeholders (donor, implementer and evaluator) starting from step one in the program design?** Beginning from the formulation of the theory of change, all stakeholders must have ongoing discussions through theory of change

then through the pre-analysis plan or logical framework of the project that are the basis for the implementation plan and the M&E plan. Conducting theory of change and the pre-analysis plan without having very active discussions between the implementer and the donor will yield little benefit. All at once, the implementer would be figuring out their logistics; they'll be figuring out their purchasing and all sorts of operational aspects, while the donor will make sure that funds are there and that their demands are met. Looking into **the evaluator side** of the relationship between the funder and the implementer, Interviewee Sixteen was able to fully illustrate the evaluator role:

> *"In hundreds of projects funded by development partners that we can use in evaluation and learn from, plenty of them had their own evaluation where they report to their own taxpayers. Thus, this made them not able to tell and criticize what is happening realistically but we were in a better position because we had a guarantee of our independence"(Interviewee Sixteen,* Director of Independent Impact Evaluation Center, May 2023).

Apparently, the best case scenario found from interviews was that the evaluator is a party that has full autonomy and is neither part of the donor nor the implementer. Taking one example about the significance of independence, Interviewee Sixteen's approach was to establish a center that has credibility and professional autonomy and they were fortunate to have an open-minded minister who agreed to keep the center independent and offered support when the center was attacked by other entities that were dissatisfied with the evaluation results. The center focused on impact evaluations and not follow-up M&E because of its limited capacity. This is due to the fact that this is the toughest type of evaluation and it takes time to see results after around six months of staff in-house training. They do not report to any development partner nor do development

partners report to them. They were able to evaluate projects from WBG, UNDP, OECD, Japanese, and German agencies (Interviewee Sixteen, 2023).

On the other hand, what normally happens is that the headquarters office of any development agency asks them to do a final evaluation of the project so they hire an evaluator to conduct an evaluation for them. The majority of Interviewees from different international organizations said that this is what happens with them whether in GIZ, USAID, UN or EU. An experienced evaluator stated that:

> *"Where the evaluation arms are embedded into implementation, obviously the evaluation reports are going to say good things about the program and it's not going to be as scientific as it should be."(Interviewee Eight, External Evaluator, April 2023)*

When the same entity that is being evaluated is also the one funding the evaluation, it becomes very difficult to obtain truthful and impartial results that provide an accurate account of the efficacy of the program.

## 6.1.2: Public Partner Involvement and Ownership

Bilateral partnerships in Egypt exhibit varying levels of engagement, commencing from the project's design phase. These levels can be categorized into three sub-levels: **active participation in designing and reviewing evaluation tools, solely approving the tool for dissemination, and no involvement in tool-sharing but participation in Key Informant Interviews (KII) for the evaluation report at the project's conclusion.** Ideally, the extent of involvement should be a focal point during the establishment of the project partnerships with public partners, particularly to ensure better understanding in the context of evaluations

(Interviewee Seven, 2023). Interviewee Fourteen was from the government side in a project related to the Ministry of Communication and she illustrated more in regards to public partner involvement saying:

> *"The public partner was the main implementer because they got the fund and they hired consultants and companies to work on that. The ministry only monitored and the donor evaluated"* (Interviewee Fourteen*, Research* Associate, National Ministry, May 2023).

However, what restricts public partner involvement when it comes to evaluations is that since the public partner is concerned with implementation, they are more focused on the monitoring plan and on data dissemination rather than the evaluation side to avoid conflict of interest.

Another example of the level of involvement when both partners are active in designing and reviewing tools is the partnership between the German agency and Ministry of Education. The German agency integrates the ministry starting from designing the program and through designing the tools. Therefore, when it comes to evaluation findings, they are able to share evaluation recommendations and support them in becoming more effective in providing the services to students as stated by the project advisor:

> *"What we do is that we share every evaluation that we conduct with our political partner, the Ministry of Education and Technical Education, and we come up with policy recommendations that they can then choose from and implement to improve their services and make them more effective"* (Interviewee Eleven, Project Advisor for the German Agency, May 2023).

This is extremely important and provides an instance in which evaluations serve their exact function: to provide ideas on how to improve projects and interventions. However, what makes this example a good one is that the implementer side was involved from the beginning

until the very end which increases their sense of ownership towards the project and they don't feel that they are just a counterpart in one phase of the service. This brings us to the final category in this sub-theme: **ownership**. An external evaluator brought up a common challenge that other Interviewees raised which is whether the indicators of the projects are internationally or nationally owned. She said:

> *"We focus on achieving the results matrix that come from the HQ, Eschborn, Zurich, DC and others but we overlook comparing the development partner indicators with our local indicators that are related to the Sustainable Development Vision for 2030. The local partners in bilateral partnerships have to have a look at project indicators and amend them so they may be integrated with the local indicators. This will heavily increase country ownership and verify that these projects are aligned with the larger local and national strategies for the country and that they are not only tied up with development partners' objectives." (Interviewee Nineteen, External Evaluator, May 2023)*

Reinforcing the significance of ownership for development projects even though they are funded by international organizations is essential to ensure the achievement of national goals for 2030. This means that the public partner has a role to play in setting their indicators, matching them with the built-in indicators of the funder organization, and monitoring and evaluating them at the end of the project. Encouraging local ownership of indicators is essential to promoting development that is both impactful and sustainable. Concepts like "use of country systems" and "local ownership" are crucial when it comes to the efficacy of aid (Holzapfel, 2016). Incorporating national standard indicators guarantees that the metrics used are culturally appropriate and represent the specific needs and priorities of the nation, while also empowering local stakeholders. This strategy fosters a sense of dedication and involvement among local

actors in addition to increasing the relevance of development initiatives, which eventually leads to more community-driven and sustainable outcomes.

### 6.1.3: Political Agenda versus Morality of Development

It is important to distinguish between research and evaluation, in other terms, scientific projects and practical projects. There are other projects **that are scientific** and evaluators have this clear separation of powers and separation of responsibilities between the three arms or the triangle of development. But as for the donor, when it comes to evaluation in practice, evaluators can't really do much. If the **donor's agenda is a political agenda**, they're not going to be interested in scientific evidence. It's not in their objective function to help evaluators achieve their own goals. If they have mutual objectives, they will help you reach there, but if it contradicts with their own objectives, then they're not going to help (Interviewee Nine, 2023). Interviewee Nine gave an example that is not necessarily about a specific international agency but in general. If an evaluator goes to an international funding organization and asks to include an evaluation that involves an RCT, the answer would be "no" because it costs money and time and is going to delay the intervention. An Intervention without an RCT is a lot faster, a lot shorter and a lot easier, so most international organizations will not be interested. In fact, an MEL Advisor agreed with this sentiment:

> *"Taking into account that if you find it more suitable to choose a cross-sectional or longitudinal approach in your evaluation that needs a quasi-experimental method, donor organizations might not approve such a method and approach. Because you are working with beneficiaries in large communities, it is hard to have a sample frame or divide the population into treatment and control groups. That is why we tend to use more flexible evaluation methods that depend on the stakeholders as well as donor requirements that*

*take care not to cause any harm to beneficiaries." (Interviewee Seven, MEL Advisor, May 2023).*

When an evaluator is asked to conduct an evaluation in the field, finding a balance is essential. It entails evaluating efficacy honestly while negotiating the requirement to communicate results in a way that is acceptable. When it comes to foreign development aid funds, particularly government-sourced funds, rigorous impact evaluations are frequently not insisted upon, especially when political agendas supersede quantifiable results. This hesitancy is explained by the significant time, cost, and complexity that come with carrying out in-depth assessments (Baker, 2000). For example, giving out EGP10 worth of rice to ten families can be completed quickly and effectively in two hours, but conducting a thorough evaluation requires a more involved process that includes gathering data on a variety of topics such as gender violence, marriage, education, and health to ensure that the intervention targets those in need of it (Interviewee Nine, 2023).

> *"If the donor's agenda is a political agenda, they're not going to be interested in scientific evidence. It's not in their objective function to help you achieve your own goals. If you have mutual objectives, they will help you reach there, but if it contradicts with their own objectives, then they're not going to help. But if it's a fund coming from a scientifically driven organization, like Bill & Melinda for example, Spencer Foundation, Ford Foundation or NSF, then these organizations will fund projects because they really want to know the effect of one intervention on a particular outcome, how it worked, how did it not work and if it worked, how it can be scaled up." (Interviewee Nine, Development Economist, May 2023)*

However, these foundations mentioned by the Interviewee are more interested in research than development projects and are not working heavily in Egypt except for J-PAL MENA (Dhaliwal,

2022). In practice, these organizations are varied in focus and vision and that is why they choose different evaluation types in different rigor scales.

However, in the development stakeholders framework, having only two players in a project is going to be a lot more efficient than having three players. This happens when implementation and evaluation are combined within one organization. **Practitioners are very keen on speeding up** the processes to show quick impact. This is despite the fact that interventions might not have any impact without taking the time needed and sticking to evaluator guidelines (Ravallion, 2009; Interviewee Fourteen, 2023). **Researchers or evaluators,** on the other hand, want to make the intervention happen slowly without any other services interfering.

For example, in one **USAID project, they planned to** do a baseline study but until they were able to do it, nine months had already passed from the start of the project. In another experience that Interviewee eight had, the baseline study was conducted a year and half from the project starting date. Sometimes the baseline takes place at the start of the project as the baseline assessment, but unfortunately sometimes baseline happens at a very late stage. A project may have only started 5-6 months ago and a baseline study may be conducted. In this case, evaluators got commissioned after implementation had already started (Interviewee Eight, 2023).

> *"Academia is more abstract than practice and there are challenges and different real scenarios that need to be addressed. In an evaluation in development, I have a lot of variables that include thousands of project inputs that I have to apply DAC criteria on in any evaluation center. While in academia, it is not the same" (Interviewee Seven, MEL Advisor for USAID funded project, May 2023).*

**One last challenge when comparing evaluators to researchers is that practitioners lack the required competencies**. Practitioners must learn many terminologies and essential

research backgrounds before working in practice. They may not know the difference between or have come across in the field the difference between 'explanatory and exploratory': if we should start with mixed methods by quantitative data first or qualitative because if we started with quantitative it is called "explanatory" and if we started a data collection with qualitative data first it is called "exploratory." What happens in the field is that both approaches take place at the same time. Between the nineteen interviews conducted, no one said they had ever heard the terms "quasi-experimental or experimental" from anyone working in the field except those with a strong research background. People call it a mixed approach, meaning they do not know the difference between research type and method (Interviewee Seven, 2023).

## 6.2: 'Impact does not happen overnight': Measuring Impact between Theory and Practice

### 6.2.1: Short Intervention Cycles

One of the critical challenges when addressing evaluation for development projects is that **the interventions are short-term** as highlighted thirty-three times by Interviewees. This dilemma was well-described by Ravallion when he elucidated the term **"myopia bias,"** where practitioners favor yielding quick results rather than long-term goals (2009, p.32). Interviewee two had similar thoughts and said:

> *"This needs to change in defining intervention results in Egypt; that impact only happens after a while. I cannot say after a very short intervention that I changed people's lives … Some of the projects have huge claims and very ambitious theories of change."* *(Interviewee Two, External Evaluator, April 2023).*

Interviewee two spotted an issue that highlights the bias for short term results: namely the notion development partner organizations have that impact can happen quickly, even though significant impact takes time. A more accurate description would be to say that it is another level of intervention, but not of impact, and so it is essential to use the correct evaluation terminologies when it comes to results to avoid oversimplification.

Then, three Interviewees thankfully illustrate what happens in practice regarding the **evaluation cycle** and the number of years each evaluation cycle must cover. Interviewee Seven clarified that:

> *"In most cases, in my experience, if the project duration is three years, the evaluation cycle is the baseline, midterm, and then end-line evaluation. If it is two years, it is only the baseline and end-line. If it is more than three years, the cycle will be longer and will*

*include an impact evaluation after phase one completion to inform the design of phase two." (Interviewee Seven, MEL Advisor for USAID funded project, May 2023)*

Understanding these cycles made it more apparent that there is no way a project that is two years or less in duration can measure long-term goals, but the impact of a project that lasts more than three years can be measured if it was planned from project inception.



Figure 5: Illustrated by author from Interviews to describe the evaluation cycle in practice.

## 6.2.2: Repetitive and Redundant Program Designs

Take two examples of how **repetitive program designs** can be helpful or not, from a refugee project design compared to technical education and dual systems projects. In Egypt's technical education and dual systems, development partner organizations have worked to construct technical education systems for the last three decades. Interviewee Nineteen is a development practitioner working in a German agency who offered a closer look at the evaluation of the technical education and dual system and said:

*"For example, in GIZ, they have been funding technical education and dual systems for almost 30 years. So that means if we are not evaluating implemented projects properly and gradually building on each phase through understanding what worked and what did not, then it is hard to know if we have been improving the whole technical education*

*system or not in the past several years." (*Interviewee Nineteen*, Development Practitioner at GIZ, May 2023)*

The bright side is that a three-decade goal is being developed by the massive ongoing effort to support Dual Education Systems with the support of the Ministry of Education and Technical Education (MoETE). Each part of the evaluation cycle takes place in the projects. They pause, reflect and learn from past processes and activities (Mohamed, 2023, p.112). Starting from the first program, which was called "The Mubarak Kohl Initiative" in 1994 (Adams, 2010), moving to the "Employment Promotion Project," "Enhancement of the Egyptian Dual System Project," and "Technical Support for the Comprehensive Technical Education Initiative with Egypt (TCTI) that the German Government funds in partnership with the MoETE. According to the Central Evaluation, the consecutive projects built a sustainable M&E system that generates periodic evidence-based data to improve trainings and match them with student and market needs (GIZ, 2021).

On the other hand, one of the clear examples of **redundant short-term programs** is the Education Assistance project funded by UNHCR in Egypt. The financing opportunity is available per school year and students must reapply yearly. Therefore, if a refugee student is in 6th grade and is not fortunate enough to get reaccepted, he/she will miss the school year until luck opens another door. The program has been operational in Egypt for 15 years but the awarding structure has never been changed so as to see if it would be more effective to support a small number of students through graduation rather than award the scholarship randomly every year. Interviewee Two described the program theory by saying:

*"For example, I will give you tuition that will help you to go to school for one year, and then you will find a job and life will change. Minimal evaluation cycles do not help anyone, and projects become redundant without changing the target audience's lives; namely projects that work for 15 years but have the same design"* *(Interviewee Two, UNHCR Program - External Evaluator, April 2023).*

Favoring shorter-term results is a bias against getting involved with long-lasting projects that can actually change people's lives. It is commonly known that governments are scrutinized by their citizens to publish where their tax money is going, which could pressure development cooperations to show off their efforts quickly (Ravallion, 2009). Their evaluation reports are published to demonstrate in tables how much money they spent in providing school grants, how many social cohesion training sessions they offered, how many employment promotion and skills training they conducted for refugees in the host country, but never how this refugee's life changed a few years after receiving the intervention.

## 6.2.3: Data Insufficiency

For any type of evaluation, there must be data to evaluate. Evaluations cannot start from scratch and collect data especially when it comes to final evaluations where implementation is already finished. Therefore, planning for evaluations starting from the beginning of the project is essential in order to plan when data collection cycles should take place and to ensure that necessary basic data is available that is needed for any evaluation that will take place. This is especially crucial in the case of more sophisticated evaluation types that will need to be decided on from the beginning as well to ensure data quality. The data limitation challenge is not only mentioned with Interviewees but also with secondary data. UN's SDG database suffers from data

limitation across countries as only 19% of the expected data is gathered which makes it challenging to report on indicators over time (Dang & Serajuddin, 2020). The Director of the Government's Independent Evaluation Center reported the same challenge:

*"One of the key issues in impact evaluation is to examine monitoring and evaluation systems. It is because of them that we can have concrete data from the field and without exception all projects we had worked on found out that M&E systems in projects lacked rigor, so we reflected on that and made recommendations, but we did not engage in changing those systems." (Interviewee Sixteen, Director of Independent Impact Evaluation Center, May 2023)*

**Strong M&E systems are no longer a luxury** to have in developmental organizations but a necessity. Having good administrative data or a good information management system is the key to producing stronger evaluation studies for agency projects. Having strong monitoring tools through the implementing partner could significantly reduce the cost of evaluations (Interviewee Four, 2023). However, this is still not what is happening in many well-known development partner organizations and in projects that are financed with billions of dollars to work with government entities. In one of the interviews mentioning work with a german agency, the interviewee said that:

*"I was shocked when I was asked when I first started as an external evaluator with a project why we had to gather monitoring data for the project or do data entry for the data of participants! This shows how behind we are!" (Interviewee Thirteen, External Evaluator, May 2023)*

Evaluators are normally commissioned to do either baseline or endline, but not to be an evaluator for the whole project from start to finish. Thus, the evaluator might face issues with gathering data if it is not gathered properly, especially in the endline. If the organization has a

69

strong data management system that includes all data of project beneficiaries and the data is inserted into a sample frame, then at least the evaluator can generate data that can be generalized within the project in any evaluation type chosen. Unfortunately, organizations do not invest in good monitoring data and techniques that will make evaluation easier and stronger. Strong monitoring will make evaluation smooth and more applicable (Interviewee Six, 2023). **One of the drawbacks of not planning for evaluations is to keep collecting data without actually knowing when and how it will be used.** This increases the burden of project targets by going through multiple data collection cycles that can be integrated into less and more unified ones that are systematic and planned from the start of the project (Interviewee Two, 2023). One more issue regarding data is undefined indicators since lack of clarity makes it challenging to collect needed information. For example, if there is no clarity on indicator definitions like "percentage of increase in the socio-economic status for refugees", it is hard to know what exactly is meant by socio-economic status and what is the method needed to measure increase. This will also require that baseline data be collected to determine what the increase is based on. There is no way to calculate the increase if no baseline study took place in the beginning of the project.

The data retrieval process is also very challenging given the fact that for an education project between the German Agency and the Ministry of Education, the evaluator had to go through the different directorates in the governments to ask for the database of the beneficiaries, which took a very long time to curate it. What made the issue more challenging was that since the service was delivered a long time ago, the contact details were no longer valid for a few of the beneficiaries. The Interviewee expressed her thoughts:

*"The quality and availability of the data was a real challenge because the partners were not really enthusiastic when it comes to sharing data in general." (Interviewee Ten, EU-funded project, M&E Manager, May 2023)*

Secondary data confirms the challenge illustrated by the Interviewees. The one major and primary challenge facing organizations conducting development monitoring and evaluation is inaccurate data and its accessibility (El Baradei et al., 2014). Therefore, a need for a strong data collection plan that serves the whole monitoring and evaluation system in international agencies is considered an essential step nowadays.

## 6.2.4: "It is all about the three-I's"

The three-I's theme is inspired by Duflo & Banerjee in their book "*Poor Economics*," where they explain the reasons that go towards ineffective aid and the traditional methods taken in development; they claim that simplistic thinking is what yields to **"ignorance, ideology and inertia"** (2011, p.3). Taking a closer look at inertia, which represents the resistance of practitioners, policymakers, and evaluators to change approaches that have proven ineffective in development, is crucial. However, the issue also extends to the subsequent "I": ignorance. Ignorance can hinder the ability to identify and understand the underlying problems, making it difficult to act upon what is observed to realistically change the prevailing ideologies regarding how change can be achieved.

Illustrating a case of **inertia**, when organizations and evaluators come together to decide to evaluate their program based on all **DAC criteria**, is an unfair act upon the program. OECD Network - Aid Effectiveness Working Group has developed six principles that developmental programs can use to evaluate themselves: "relevance, coherence, effectiveness, efficiency, impact

and sustainability" (OECD, 1991). The assessment framework of the six criteria was founded to evaluate any intervention in development. A.O. is an external evaluator with eight years of experience working with several agencies and development partners, mainly in the last few years with GIZ, USAID, BPRM and UNICEF. He contributed in his interview to an overlooked challenge between the six DAC criteria and the inertia dilemma.

> *"Organizations are forcing evaluators to measure impact and sustainability through evaluations, and most of the time, the projects are not eligible to be evaluated by these criteria because the projects were not designed to foresee these two measures. They list all DAC criteria, for example, without filtering from the criteria what is convenient to their evaluation questions and data." (Interviewee Eight, External Evaluator, April 2023).*

The evaluator clarified a fundamental yet passed-over idea: not every assessed intervention is designed to be evaluated by the six principles nor will it be adequately evaluated by them. The appropriate method for evaluation depends on the evaluation purpose, questions and the project nature that sometimes is the frame in which both evaluation questions and purpose work. Therefore, an intervention that is one year can not be asked by its owner agency to evaluate its lasting goals because realistically there aren't any. Impact and sustainability need long-term activities and theories that yield an effect beyond outcomes and outputs. Also, a project that needs to be monitoring its budget or has not done a cost-benefit analysis is prohibited from measuring its efficiency (Bertrand et al., 2009). Efficiency is measuring if the resources are well-spent, and this is hard to measure by traditional qualitative methods and requires a structured cost-benefit analysis. Choosing the convenient criteria and deciding what fits is outside the M&E culture in Egypt, which is a challenge bound by ignorance and inertia.

The second dilemma is about **ignorance combined with embedded ideologies** that have been cultivated and nurtured over the years, mainly the notion that aid is working very well, and that this is what must be reported to ensure funding continuity. **In a culture that does not accept mistakes**, taxpayers give their money to the government and expect swift and profound results. Development organizations expect reports from their country offices and implementing partners stating that the project was successful. This may be a reason why project managers avoid being realistic or avoid acknowledging the fact that their efforts may not be as useful as they could be to the beneficiaries. Interviewee Seventeen is a seasoned staff member with 5 years of experience working in a multilateral organization in a project funded by a UN agency; she works in an independent M&E team writing a report that is to be published. She states:

> *"At one point, my boss told me about the comments she kept getting from the middleman reviewing our work: "Oh, remove this because it is so negative! Oh, you know, do not mention certain stuff! Try to make this look better." To me, it just seems like they want to show that they are doing their best efforts and trying to conceal the challenges, and I do not think that is fair or right." (*Interviewee Seventeen*, Multilateral Organization, M&E Associate, April 2023)*

The M&E associate was shocked by the fact that their M&E department was not as independent as it claimed. The ideology of staff members that mistakes should never be known is a bias that makes the evaluation process no longer viable. Where the **evaluation arms are embedded into the system and implementation** and the organization as a whole for many reasons, the evaluation reports will say good things about the program. However, they should be more scientific. It should be acceptable and common knowledge that the number of successful projects in comparison to unsuccessful ones would be very low. So if we have ten projects, we

should acknowledge the fact that it is likely that two or three of them would be complete failures, two will be successful, and the remaining will be average. When speaking with individuals who work in the development sector, you are likely to encounter individuals who do not accept or appreciate the idea of making mistakes (Interviewee Four, 2023). To rectify this, there must be a growing ideology of bravery, wherein there is a willingness to acknowledge when something in the implementation process is not working well, to measure it, and work on changing the design. Unwavering dedication to theories, efforts, and resources while ensuring fair results must be the status quo.

**Ideologies are the driving force towards stronger theories and interventions** that serve to support development dilemmas. Michael Kremer established "Development Innovation Venture" to introduce agencies to the idea that investing in innovation is worth it. Not only that but Michael demonstrated that using or supporting projects based on scientific evidence would maximize the impact of spending money up to 5 times (Kremer et al., 2021). Many organizations established innovation ventures but few evaluate their impact thoroughly and measure their return on investment (Kremer et al., 2021). Interviewee Four is the L&D Director in a National Foundation in Egypt who also has more than 10 years of experience in various international development organizations. He expressed his perception of the importance of evaluation and its correlation with ideologies:

> *"Some of the development practitioners I worked with do not understand the possible return value and the importance of undergoing an impact study if the project is eligible for it." (Interviewee Four, Learning and Strategy Director in National Foundation, April 2023)*

Underpinning the first notion of how ideologies in themselves can be a challenge and the ways to overcome this was an eye-opening finding in the research. The finding correlates strongly with similar challenges illustrated by the way many illogical and futile behaviors can result from firmly held ideological convictions. In order to combat poverty more effectively, it is urged to critically assess these ideas and adopt a more evidence-based strategy in interventions for international development (Karlan & Appel, 2012). This perspective of the analysis proposes that perhaps it is the issue of biased ideologies.

Viewed from another angle, we face a unique problem that is not so much ideological as it is the result of ignorance or a conscious decision to continue with antiquated, poorly evaluated program designs just because they suit the donor government's tastes. Interviewee Nine, a development economics doctoral candidate with expertise in health and education program evaluations, echoes the opinions of multiple interviewees:

*"There are projects that are done for the sake of politics where the donor basically doesn't really care about the evaluation nor the impact of these interventions. All they care about is less refugees knocking on their door, or less crime happening within their border or less illegal migration happening across the Mediterranean. But they don't really care about the other possible impacts of the interventions." (Interviewee Nine, Development Economist, May 2023)*

It is true that one common philosophy that comes up often in interviews is the politicization of initiatives. In reaction to this idea, others argue that although donor governments might not have an incentive to see a return on their funding, it is still imperative to show an impact particularly when considering the viewpoint of taxpayers. It could be necessary to present concrete proof of the impact rather than mere data on the quantity of services rendered.

Additionally, practitioners should put pressure on donors to act on their suggestions for improving the efficiency of healthcare, education, and poverty alleviation programmes. Furthermore, beneficiaries may speak up for improved service quality, which would force project managers to modify their tactics. Both funders and beneficiaries themselves have the potential to exert influence.

Incentivizing evaluations can be conducted through multiple ways as previously stated. The donor could answer to some sort of accountability measures. For example, if they are getting their money from the EU, they must answer to the taxpayers and the European Parliament in Brussels. In the case of the GIZ, they must report to the ministry of development cooperation and to the German taxpayer. The entity must create an incentive to make the taxpayers want to pay. At the end of the day, the way GIZ gets its money is from funds designated as foreign aid coming from the German Government, and the government gets money through taxes, same goes with USAID. With more than 20-years of experience working qualitatively and quantitatively in research and evaluation of development programs with a more focus on the government side, Interviewee Fourteen, shared a similar point of view to Interviewee Nine:

> *"Donors want to publish in front of the congress or EU their achievements in developing countries and how much they contributed and to defend how and why they spent the money. Effectiveness is an entirely different story." (*Interviewee Fourteen*, Research* Associate, National Ministry, May 2023)

As long as the incentives go towards showing-off results, it is hard to convince international organizations funded by their countries to be more effective and mindful towards where they put their money and to consider more effective intervention ideas that can be

data-driven, more successful, and cost-effective. This is where the dilemma starts and ends about how ideologies are changing everything.

Taking an intriguing Egyptian case study into consideration, we examine **the EGRA project in Egypt**, which was initially funded by USAID from 2013 to 2020 (Stern et al., 2016) and later supported by the Literate Village Program by Save the Children International from 2020 onward (Egypt Literate Village, Early Grade Reading Assessment (EGRA) Baseline Report, 2022). The baseline report for this project was completed at the end of 2022. Interviews with Interviewee Seven from the USAID team and fourteen from the Ministry of Education shed light on the diverse range of ideologies that come into play when it comes to evaluation methods and decision-making processes:

> *"In the education project commissioned by USAID - (EGRA) the evaluation depended on intervention vs. control group which is an experimental way of approaching it and this is because the tool was a very scientific and universal education tool. The USAID-Egypt argument in this project was ethical because you are deciding to prevent students from taking the intervention compared to other students which is unethical. In social science there is a huge critique of experimental methods. This is unfair in my opinion. If you choose to use non-experimental and have a baseline against endline study you will understand the reading capacities and skills before and after the intervention."* *(Interviewee Seven, MEL Advisor for USAID funded project, May 2023).*

Interviewee Fourteen disagreed with the aforementioned perspective from a scientific standpoint. She contended that there are requirements that must be fulfilled in order to apply pre-and post evaluations. Among these is making sure the sample's economic circumstances don't change. Even if their conditions get better, it might not be because of the intervention; other variables like changes in consumption habits or the death of a family member can also affect

results. According to Interviewee Fourteen, evaluation studies may yield dubious conclusions since practitioners frequently employ pre-and post-assessments without properly taking these outside influences into account (May, 2023). While the assessment tool has its own requirements when it comes to evaluation, ideologies of implementers and development organizations can have different opinions about the standards when it comes to different contexts. The Ministry's researcher had another attempt with EducateMe and RISE Egypt to evaluate the EGRA as it should be with both pre and post as well as before and after to ensure getting the most accurate results.

## 6.3: Purpose-related Challenges and Opportunities

This theme is mainly focusing on understanding the different types of evaluations and the prerequisites for each type or purpose of evaluation to ensure conducting rigorous evaluation rather than just a tick-the-box evaluation.

### 6.3.1: "Tick the Box" vs. Comprehensive Evaluation

This section starts with dividing evaluations mentioned by Interviewees in bilateral and multilateral partnerships according to the purpose of conducting the evaluation study then going through what strong evaluations should look like and why it is recommended to do one.

The next few quotes represent the perceptions of the interviewees on how they think development partner organizations conduct evaluations - this part of the questionnaire was inspired by the literature review. **Liket surveyed** more than 30 organizations funded by donor governments about their purpose for conducting evaluations. The results revealed that 39% pursue evaluations to **check the box of the funder requirements** (2014). The M&E Associate

who is also a member of the Egyptian Evaluation Association was asked why she thinks bilateral partnerships undergo evaluations:

*"So I think that the main reason they undergo evaluations is unfortunately to simply **tick the box**. I don't think all organizations are like this, but some are unfortunately becoming donor-oriented. This means that they conduct an evaluation because the donor asked for one and not because they personally wish to do so. This doesn't reflect a positive mindset."* (Interviewee Seventeen, *Multilateral Organization M&E Associate, April 2023)*

It seems not only to her but to the majority of interviewees that this is the case. The headquarters of all the projects require that they submit a report indicating that they underwent a final evaluation of the project to be able to finalize the project. It was surprising to also find an external evaluator expressing similar sentiments when asked the same question:

*"Organizations undergo evaluations because they have to do it, as was the case with 70% of the projects I worked on - which were with different development partners in more than 10 projects. They do evaluations because it is a requirement to do them whether baseline study, gender analysis or endline. For example, once clearly in a very straightforward way I was asked to do an evaluation in 10 days because it was a requirement and they needed to submit it! Can you imagine? And this example was from a German Organization funded project"* (Interviewee Eight., *External Evaluator, April 2023).*

Tick-the-box evaluations do not consider whether the measure or intervention was impactful or not. It is treated as a **checkpoint that must be passed** by any means. This makes the evaluation useless and empties the study from its core purpose. In the next quote, the generated idea from analysis is further confirmed by W.S. who is an advisor who has been working for GIZ for almost 15 years and describes when evaluations take place in the project cycle:

*"Normally, we do the final evaluation with the new phase appraisal combined to save costs." (Interviewee Eighteen, Senior Advisor in a German Organization, May 2023).*

This discovery highlights an unexpected fact: the project's next phase begins regardless of the evaluation's outcome, suggesting that the outcome of the final assessment may not really matter. Given this revelation, one could wonder why evaluations are conducted in the first place. Evaluations seem to take place more so to verify that the organization is prepared to move forward with a new phase, **for the same project or one comparable to it, than to judge if the project was successful or unsuccessful.** This is exactly what tick-the-box evaluation means. It is not necessarily the only type of evaluation that one organization adopts; it varies inside the same development partner organization as per the interviewees in the five agencies we are focusing on. But surely this type of evaluation means that whoever does it considers a development project as a business project that is continuing indefinitely without the potential to be enhanced or improved based on data.

The following is a list curated from all the interviewees on what they consider to be a rigorous evaluation when it comes to development partner organizations in Egypt:

- **detailed evaluation plan, (cost challenge)**

- **timing of conducting evaluation,**

- **generalization conditions,**

- **scientific sampling technique,**

- **robust tools that have been tested,**

- **report writing rules.**

Each point will be further illustrated in the following paragraphs.

Following the literature review, **the absence of evaluation planning (cost challenge)** increased the missed opportunities for improvement and knowledge-sharing results from evaluations (Toscano, 2013). Evaluations are either not planned or are conducted with low quality as a sense of a duty to fulfill, which leaves the task's core aim unfulfilled. Holzapfel introduced the evolution of the RBM framework in the 1980s by UN agencies as a suggestion by the OECD to increase M&E effectiveness. The framework continued to evolve over 10-15 years to support agencies in building their short and long-term project objectives (Thomas et al., 2021; Hoefer, 2000). RBM's primary purpose was to increase aid effectiveness in development cooperation. The framework is used mainly to plan for the whole Monitoring and Evaluation plan as it tracks the progress of activities, indicators and objectives on all levels. Each indicator has its own method of monitoring its progress and measuring it with the frequency suggested and planned prior to the start of the project (Holzapfel, 2016). In practice and as described by practitioners questioned, evaluations are considered in the project budget at the planning phase by normally estimated "EUR 20,000" (Interviewee Thirteen, 2023) then planned thoroughly after implementation starts.

> *"The absence of an evaluation plan during the design phase negatively impacts the evaluation process, preventing the comprehensive capture of the project's true objectives and hindering the attainment of a meaningful evaluation." (Interviewee Two., External Evaluator, April 2023).*

**The second challenge** is the **timing of conducting evaluation**, which is critical to ensure accurate measurement. Therefore, as stated, planning for an evaluation should happen before implementation starts in the planning phase as the duration of each evaluation cycle is important

(Thomas et al., 2021). Interviewee Fourteen pointed out a rule that she learned was very important in one of her experiences in an international cooperation project:

> *"The challenge in evaluating interventions in community schools arises from a lack of understanding among practitioners and stakeholders that concurrent interventions should not be implemented and evaluated simultaneously." (*Interviewee Fourteen*, Research Associate, National Ministry, May 2023)*

What happens in practice in development projects is that if a project is three years long, planning and assessing the project takes place at the end of the second year to know results by the end of the three years. But in this case we cannot actually say if the community was impacted because there must be a time gap between the intervention and assessing the outcome. Confounding outcomes might arise from practitioners and stakeholders not understanding how important it is to avoid implementing and evaluating several interventions at the same time. Conducting evaluations in such a manner may make it difficult to ascribe observed effects precisely to one intervention, therefore undermining the reliability and effectiveness of the evaluation process.

After planning an evaluation and choosing when it can take place comes the **selection scientific sampling technique.** The data that evaluators come across after implementation necessitates the use of an appropriate sample technique. Scientific sampling procedures must be used during sample selection if the objective is to generalize findings to the target group, regardless of counterfactual or intervention contribution and attribution issues (Interviewee Eight, 2023). If the chosen sample is too small or too big each has its own drawbacks. The adequate sample percentage that is based on the choice of confidence level and margin or error is the ultimate solution. Traditionally, evaluators used to select the technique and percentage

without a scientific background (Toscano, 2013), given the fact that it is not necessarily that the technical expertise for a development practitioner means they have research experience and background. It was known that 10% sample is the fixed and common percentage to take from a sample no matter how large or small the population is. When we start to seek consultation of experienced researchers who know scientific methods and techniques for samples, they change this misconception to a formula that we run to know the sample percentage that we need to take to say this is a representative sample from the population. However, there is a sample size calculator and six-different types of samples that are chosen based on the evaluation purpose (*Sample Size Calculator*, 2023). Interviewee Seven, narrated a challenge that happened to her once in a project:

> *"I worked for a project population of 20,000 beneficiaries in four governorates. The baseline study should not be more than 100.000 EGP… understanding the essential research knowledge gave the project the know-how to better utilize the financial resources" (*Interviewee Seven*, MEL Advisor for USAID funded project, May 2023)*

Her example describes not only the barrier for practitioners to know how to have a good sample but also a situation in which **the resources used for the study are limited** and the majority of indicators are quantitative. In such a case opting for a purely quantitative approach rather than a mixed methodology could significantly compromise the quality of the evaluation. By selecting a smaller sample size, the confidence level of the study gets reduced. A follow-up to this challenge is choosing the most convenient methodology for the study.

**A tight methodology** is triangulated and piloted among different people to ensure questions are understandable as designed, that tool is based on the local context and not taken copied from the HQ since this might not fit the culture that is going to use the tool.

Another aspect affecting analysis and report writing has to do with a typical reporting mistake related to the methodologies used, whether they are qualitative or quantitative. For example, a significant error happening in practice but not mentioned in secondary data in **qualitative analysis is to rely just on notes**, to ignore transcription, and to choose an erroneous scientific summary of the notes. Interviewee Four, illustrated that by saying:

*"A lot of non-experts that perform evaluations rely on qualitative instruments, which might result in inaccurate and biased reporting, because assessors may include their personal opinions without deep analysis that is required for qualitative methodology. More than 90% of the studies that I have seen in the last ten years have used qualitative research." (*Interviewee Four*, Learning and Strategy Director in National Foundation, April 2023)*

Non-research background practitioners tend to find qualitative easier than using quantitative since it is based on interviews and written notes. However, scientifically, both methodologies have their own complexities and the hypothesis that qualitative does not require much work might indicate that the practitioners do not know how to analyze qualitative data. It is not only qualitative methodology that has its own challenges but also mixed methods approach as well. Interviewees in mixed methods identified a common mistake made by practitioners who divide up qualitative and quantitative studies into discrete sections, ignoring the triangulation system designed to combine results from both approaches in a seamless segment.

## 6.3.2: The Dilemma Around Proving Causality

Causality relationship is the final characteristic among the six-principles to mark an evaluation as comprehensive and strong. However, proving causality is not as it sounds because each intervention is surrounded by external factors that make it hard to isolate the beneficiary

from it to say that the intervention was the only factor that led to better economic circumstances or health conditions for the beneficiary for example. Otherwise, there might be other factors that led to these better conditions. To qualify evaluation to be rigorous or not based on causality is an ongoing debate between practitioners. One side says:

> *"Rigorous for me means any tool that can prove causality. The well-known tool for proving causality is Randomized Control Trials (RCTs). I will be very hesitant about any other tool. For me, non-rigorous methods mean qualitative tools"* (Interviewee Four, *Learning and Strategy Director in a National Foundation, April 2023).*

Randomised Controlled Trials (RCTs) are widely regarded as the most scientifically robust approach, with other methods regarded as comparatively less accurate. RCT is a natural experiment that basically involves no intentional randomization of treatment assignment for the purpose of study (Koolwal et al., 2010). For example, if the government decides to provide health insurance and they ask 100 participants to apply they say: "we have money for 100 participants." If there are about 10,000 people applying, then they say: "we're going to do a lottery and the result of that lottery will be the people who get that money" (Interviewee Nine, 2023). RCTs should be conducted if financial and resource constraints allow. Prioritizing RCTs is also recommended if one has influential connections, collaborates with partners, or has an exceptionally innovative idea. In cases where these factors do not apply, regression analyses such as difference-in-differences or other rigorous tools can be used. In the event of additional constraints, simpler methods such as Simple Difference or Pre and Post can be considered, particularly in qualitative research (Interviewee Four and Fourteen, 2023). This approach is guided by the principle of beginning evaluations with the most rigorous methods and then exploring alternative methodologies if these are impractical.

When using pre and post measurements in quasi-experimental or non-experimental settings, there are a lot of assumptions involved, especially when it comes to outside variables like devaluation. The same population that is facing difficulties is affected by devaluation, which results in lower rather than higher consumption or maybe someone in the family died and that is why they were able to save money better than before and not because of the intervention. These factors are hard to measure but they interfere in evaluation results. Pre- and post-measurement conditions should ideally be similar, but this is frequently not the case.

> *"Traditionally, evaluation in practice was providing services and coming afterwards to see how it was and take feedback. There is no condition to do a baseline evaluation but they might do a retrospective evaluation by asking how much your salary was before and after taking the intervention but this is very inaccurate and doesn't show the effectiveness of the service."* *(Interviewee Fourteen., Research Associate, National Ministry, May 2023)*

Another case study that illustrates the same ongoing idea is the impact evaluation used in a well-known project at the German Agency where they provide career guidance sessions to students to help them reach better employment opportunities. The evaluation conducted took place before and after to assess the project's impact with the help of the Ministry of Education; they had two groups of students who took the sessions and another group with the same characteristics who did not. It was a retrospective evaluation. The staff member associated with that evaluation described the reason that motivated them to conduct an impact evaluation:

> *"In the case of the impact evaluation that I conducted with the control groups, the indicator mentioned that "two impact evaluations using control groups are implemented". There I knew I had to do quasi-experimental design with treatment and control group"* *(Interviewee Ten, EU-funded project, M&E Manager, May 2023).*

Retrospective evaluations face a number of challenges that limit their effectiveness. The reliance on historical records and the scarcity of data make it difficult to gain a comprehensive understanding of the project's dynamics. Memory bias raises the possibility of distorted recollections, jeopardizing the accuracy of information provided by participants and stakeholders. Furthermore, retrospective evaluations struggle to address new issues or changes in the external environment that occur after the project is completed. Causation becomes difficult to establish, making it difficult to determine whether observed outcomes are directly attributed to the project or influenced by external factors. The evaluation's comprehensiveness is further constrained by limited stakeholder involvement as individuals may no longer be actively engaged.

The other side of practitioners see that as beneficial as RCTs may be and though they may try to measure what would have been if the counterfactual can be created, the problem is in the context. In addition to that, splitting people into two groups, the intervention and the comparison, is through a lottery. This side of practitioners see that this setup does not fit in the Egyptian context. That's why although it may seem like a good idea on the outside, on the inside you have to consider the context that you're in. Targeting the most vulnerable people ever and telling them some are taking money and some are not, this does not fit in the Egyptian context (Interviewee Seventeen, 2023).

Through deep analysis, the answer to having rigorous evaluation does not rely on one side rather than the other but on understanding that in order to conduct some experimental types of evaluation then certain conditions must be fulfilled. If those conditions are satisfied, then it is necessary to go for RCTs, but if they are not then it is advisable to resort to other rigorous

methodologies. Choosing a project for impact evaluation entails reviewing its documentation and critiquing the general terminology used. The process starts with defining specific goals, which is especially important when development partners use broad language like "improving public health in population x" which necessitates a precise definition of each term and the delineation of operational pathways. It is critical to evaluate the project's Monitoring and Evaluation (M&E) framework and question if it includes periodic progress reports and a mechanism for adapting the original plan based on findings or not. The availability of pre-existing data is required for an impactful evaluation; inventing data that is not already in place undermines the integrity of the evaluation process. Overall, taking these criteria into account ensures a more meaningful and methodologically sound project evaluation (Interviewee Sixteen, May 2023).

### 6.3.3: Purpose-based Evaluation Challenge

After considering the different motives behind choosing to conduct an evaluation, it is now possible to merge them into four main reasons: insight provision, decision making, development and learning, and policy making (El Baradei et al., 2014; Interviewee Eight, 2023; Liket et al., 2014; Rist & Morra Imas, 2009). As expressed in the perspectives shared, learning emerges as a critical purpose of evaluations. The emphasis on cost-effectiveness highlights the importance of evaluations in optimizing resource utilization in the face of constraints. Evaluations help to determine when investments should be prioritized and
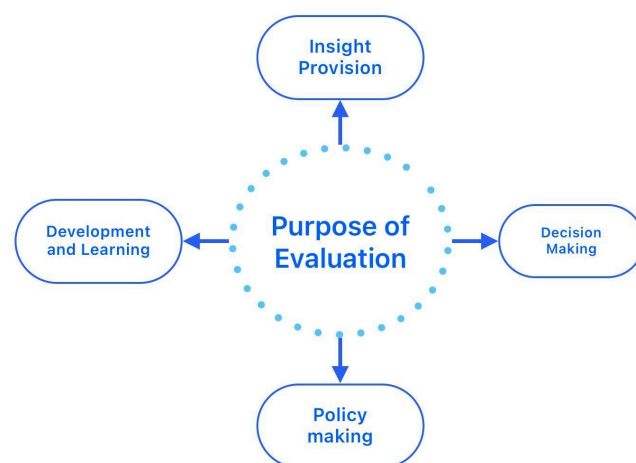
when prudent resource allocation is required. Others, however, argue that evaluations are necessary for the project's inherent learning process. Evaluations go beyond meeting indicators to comprehending project dynamics: identifying what works well, what needs to be improved, and gaining insights into how beneficiaries perceive project initiatives. In essence, evaluations contribute to the overall effectiveness and impact of projects by serving the dual purposes of cost-effectiveness and continuous learning (El Baradei et al., 2014). In addition to that, Interviewee Seventeen had another perspective that is related to evaluation learning:

> *"We conduct evaluations to learn about the best practices and the worst practices. We also conduct them to 'unlearn' which is a great term that means not doing the same things just because they work." (Interviewee Seventeen, Multilateral Organization M&E Associate, April 2023)*

Figure 6: Illustrated by author from Interviews to describe Purpose of Evaluations

The ideology of learning and unlearning practices is new in the Egyptian context. Although the majority of practitioners and project managers are open to conduct evaluations sas long as they are a requirement, they may be rigid and may hold on to some practices that are old-fashioned or that data has proven to be wrong. Hence, learning and development play a crucial role in changing the traditional way of doing things.

After learning what works and what does not work, the next step is taking decisions based on the data derived from evaluations. This aspect will yield great results if decision makers are convinced that decisions should be made based on evaluations in addition to any other factor they might want to take into account. The concept of evaluation is critical for those who truly believe in the development process. By recognising our basic presumptions, decision-makers should use thorough tools for evaluation to objectively determine what is and is not effective in

the context of initiatives for development. Evidence-based proposals through including evaluation and needs assessments to support new projects or phases of existing projects. In order to improve the reliability and efficacy of the proposed initiatives, we must propose new designs based on evidence and data.

> *"On a bilateral level, when project managers see the added value of M&E functionalities in their project, they start seeing why it is important and they move towards embedding the M&E system into their projects in different partnerships and involve their public partner in it." (Interviewee Eight, External Evaluator, April 2023)*

Through prior evaluations, it was possible to analyze to what extent project managers have the power to inspire both the public partner and the implementers to take their decisions collectively in the partnership. Donors who fail to engage with implementers and partners frequently neglect their participation and involvement approach in evaluations. This omission has a major effect on the evaluation's general quality. For example, in a project partnered with the Ministry of Education, beneficiaries and public officials who were part of the project in its first year were interviewed after five years. When this was done, it became clear that their data was outdated in the development agency database and whoever was still there had mostly forgotten about the intervention (Interviewee Nineteen, 2023).

Bilateral projects mainly work towards implementing Egypt's vision 2023. In this regard, new projects are mainly designed to support the different ministries in implementing their objectives towards the country's main vision. This was acknowledged by one of the Interviewees whose work in different impact evaluations was mainly and purposely done to give insights to the minister's decision and future work:

*"Our evaluations are mainly about providing the minister with insights on various projects as Egypt receives $1 billion in foreign funds each year in a variety of sectors, including health, water, industrial, agriculture, and education"(Interviewee Sixteen,* Director of Independent Impact Evaluation Center, May 2023).

We can design interventions that are evidence-based by looking for observations from those with first-hand knowledge and using data to gain an understanding of programme outcomes. This approach not only improves project effectiveness but also enables expansion, guaranteeing that effective approaches can be implemented in broader contexts.

Legitimation entails the process of converting the results of projects into case studies which act as informative lessons for future initiatives as well as generalizing the findings to the policy level. This approach aims to encourage and motivate policymakers by showing the project findings' relevance and significance in more expansive regulations settings. According to Elbaradei et al., only 41% of evaluations performed in organizations for policy level purposes compared to 74% for project level (2014, p.14). The percentage emphasizes the urgency to engage in policy-level and purposeful legitimation evaluations.

*"It is not eligible to be generalized because normally legitimation happens when we do impact evaluation to the projects that we can generalize its data. (Interviewee Eight, External Evaluator, April 2023)*

Impact evaluations are the only road when it comes to legitimations. Since impact evaluations are hard to navigate in all project settings that consider costs, it may be helpful to identify stakeholders and entities involved and collaborate effectively with them. Aligning all these factors altogether to not only evaluate the project on outcome and output level but also impact requires the six-conditions altogether that were previously mentioned in this section. On

that basis, conducting evaluations with the aim to inspire or change policies is considered rare in Egypt's bilateral and multilateral organizations. From a partner level, they must endorse evaluation results, which means the partner is able to share the results at the strategic level and it can be referred to when designing policies at the country level or when scaling up in a similar project to take it as a case study.

## 6.4: Evaluation Utilization

In order to improve the usefulness of evaluations, Patton suggests examining the obstacles preventing their efficient use (1978). These issues can be summed up in the findings as follows: there is a lack of data, stakeholders are not very interested in evaluations, and there are not enough learning resources and methods that can be used for evaluations. Evaluation is also utilized for institutional reforms, system development, piloting to come up with good practices and tools.

### 6.4.1: Urgency of Stakeholders' Engagement in Evaluation

One of the main success takeaways related to both implementation and evaluation is using a **participatory approach** in evaluations when needed. Including the main stakeholders in the evaluation process yields a better understanding of the knowledge and stakeholders are able to suggest better recommendations for the services provided (Sette, 2021). This approach automatically makes the main project stakeholders understand the significance of monitoring and evaluation and how their words will greatly impact the intervention which leads to creating a common understanding culture of M&E. As part of learning from past experiences, taking education reform 2.0 as an example, Interviewee Fourteen further illustrated the issue:

> *"The curriculum in K-12 schools was changed but they did not train the teachers on these changes and the new curriculum which made the whole system fail. Considering all stakeholders is important and is a challenge" (Interviewee Fourteen, Research Associate, National Ministry, May 2023).*

After digging deep into the new reform, it became apparent that the curriculum was only introduced to centralized trainers and not front-liner teachers who were going to deliver the

curriculum. Not only that, there was no orientation for parents except from the media which left a gap filled with rumors that did not help the reform and led it to failure (Zahran, 2023). These challenges would have not happened if the frontline teachers, parents and students were included in the discussion as main stakeholders for the projects. Facilitating an ongoing feedback pathway between stakeholders is what can make new innovations succeed and support the whole development project (Toscano, 2013). The silver lining is that one of the practitioners was able to identify that solution and is encouraging other organizations to undergo more frequent qualitative research to better understand the different target groups and what they really need. As a bilateral partner, the organization should not avoid to talk to any partner except their political partner. According to Interviewee Ten, they should talk to beneficiaries as well:

> *"It's only when I understand their fears, dreams and needs that I can really design interventions that serve them in a good way. That's why I feel that focus group discussions are something that we can easily do ourselves as we do not necessarily have to always outsource these things. Through our M&E staff in the projects, we can simply do a focus group discussion from time to time" (Interviewee Ten, EU-funded project, M&E Manager, May 2023).*

Keeping an open communication channel with stakeholders along the project duration by inviting them to different events and keeping in contact with the partners is a good practice but not a very common one, unfortunately. The myth that involving participatory approaches will waste time is not really true as in return the quality of the evaluations increases dramatically.

## 6.4.2: Learning Mechanisms Through Evaluation

In the last sub-theme in the analysis, the focus is turned to **how much we can learn from evaluations, and how to document these processes and methods to disseminate knowledge**.

The participatory approach in evaluations inspires the project mindset to be more humble towards gaining knowledge from the field and beneficiaries as well as from the counterpart and development organization team (Baker, 2000). This changing mindset will allow us to go the extra mile and understand that maybe the intervention requires checking on an aspect that was overlooked while designing the project. In a UN education project, Interviewee six gave an example that illustrates the idea very well. They implemented an education project for the community and at the end of the Focus Group Discussion when instructors asked if anyone had something to add, a group of beneficiaries said that they can not access the school to participate in the activities there because there is a risk while going there. This was also affecting the results of the data collection tools gathered and the community attendance in the FGDs. Such a challenge was not part of the services offered but a protection issue.

> *"Evaluation is not only about working on results and findings but also extracting unlearned factors and services that need to be offered to the community" (Interviewee Six, M&E Manager, UN Agency, May 2023).*

This finding from qualitative research would have never been discovered if the research were quantitative. As suggested by a few of the interviewees, it is important to encourage international cooperations to understand the beneficiaries' backgrounds in order to better design and tailor the services offered to them. This is essential when it comes to evaluations and in specific participatory approaches to serve the utilization of evaluations. The approach not only serves changing factors in interventions but changing the project as a whole when the staff learns things that may change the project design. In a women-empowerment project that is implemented in rural areas, the national foundation was able to shift the decision-makers'

attention to an overlooked factor based on evaluation data of a previous phase. The Interviewee described this incident:

> *"Because of how social norms affected our last project, the new project has a huge component that focuses on tackling negative social norms. It focuses on how we can introduce responsibilities for wives and husbands in households and how we can support women to positively interact with the family and also how a woman can negotiate with her mother-in-law" (Interviewee Four, Learning and Strategy Director in National Foundation, April 2023).*

The social norm talked about in the quote was how mothers-in-law are the household decision-makers between the wife and her husband especially in rural areas in Egypt. The national foundation decided not to overlook such a finding and integrated inviting the mother-in-law in the training of the women. Such a learning point helped increase attendance rates, changing in mindsets for women that served as part of the project outputs and outcomes.

When talking about documentation of evaluations, there are **two challenges that hinder it. The first is related to choosing where to allocate this information and how to retrieve it, and the second to who should be responsible for this knowledge.** In organizations, normally, the knowledge is there but there are no proper communication channels. For example, if a project ends and the evaluation report is available, and there is no team working on the project anymore, who should be responsible for disseminating the findings? Is it the person who will write a new proposal on the same area of the past project? Or is the M&E unit responsible for that? The question is not only concerned with disseminating but also with making sure that they are utilized properly. From the organization's side, there must be emphasis on asking evaluators to prepare innovative methods on how their report is going to be disseminated. This should be

required in the Terms of Reference for the evaluator, rather than simply asking them to submit the report. Dissemination should be conducted in the easiest and most accessible way to the stakeholders. During the research efforts to observe the evaluation scene in Egypt for the collection of data for this thesis, USAID held an event to share their baseline findings in a project about women empowerment which was a good practice for cross-organizational communication and learning from each other. Oxfam also had a similar event but for the final evaluation. They organized a learning event to share evaluation findings and this opened a discussion not only about their project and organization but also for all development practitioners working in the same sector and are interested to learn from their experience in the project. Aside from events, most development organizations have an internal library where they store all project documents but it is unclear if it is used when designing new projects or not.

## *Chapter Seven: Study Conclusion and the Way Forward*

### 7.1: Study Conclusion

Finally, with an emphasis on Egypt specifically, the research began a thorough investigation of evaluations carried out in bilateral and multilateral development organizations within the framework of projects, programmes, and funding from the UN, German agencies, and American agencies. The study began by breaking down the connections between the results framework and the development goals, highlighting the importance of relevant, effective, efficient, and long-lasting international development interventions. The study revealed differences between tick-the-box evaluations and rigorous evaluations using a qualitative methodology that included 19 Key Informant Interviews (KIIs) with practitioners, researchers, agency staff members, decision-makers, and government officials. It emphasized the difficulties in assessing development partnership projects in Egypt and the opportunities lost because thorough evaluations were not conducted.

The findings carefully analyzed the current state of evaluations, approaches to efficient evaluation, and evidence-based recommendations for evaluating development projects in Egypt. This addressed the main research question regarding the opportunities and challenges facing evaluations in donor-funded development projects in collaboration with the Egyptian Government. The conceptual framework outlined the intricacy of evaluation types, such as monitoring, process evaluation, cost-benefit evaluation, and impact evaluation. It was based on findings from fieldwork and a literature review and asserted that rigorous evaluations result from the complementary use of multiple methods.

The study's findings, which emphasize the need for a common language for evaluation practitioners, increased public partner ownership, the creation of independent, multi-layered evaluation institutions, and the growth of evaluation stakeholders' capacities, provide insightful information about the evaluation environment of development projects in Egypt. Project managers, implementers, public institutions, M&E staff, evaluators, and academics working on bilateral and multilateral projects can use the findings as a guide to help build a stronger and more efficient evaluation culture in the context of international development interventions in Egypt.

## 7.2: Recommendations and Way Forward

The section on recommendations seeks to address critical issues raised by the research questions. It focuses on Egypt's evaluation context, ways of enhancing evaluation procedures, and recommendations based on solid evidence for bolstering evaluation systems in development partner organizations. The recommendations, based on the study's findings directly, will investigate novel evaluation methods while highlighting uncharted territory and suggesting creative solutions. Project managers, practitioners, those in various levels of monitoring and evaluation roles, evaluators, and academics involved in bilateral and multilateral projects are the specific targets of these recommendations. The goal is to offer practical insights that advance the field of evaluations in the context of international partnerships and Egyptian public organizations by connecting these recommendations with the research questions.

### 7.2.1: Unifying Evaluation Terminologies within different organizations

Acknowledging that there is a misconception in different terminologies is the first step. For example, it is important to know the difference between project evaluation and impact evaluation because they are not synonyms. The conceptual framework of this thesis is to shed light on the wide-variation in terminology and to warn against using terms and concepts interchangeably as this can lead to stakeholders thinking that impact evaluations are the only type of evaluations and many more myths. Hence, we are calling for a dictionary to be prepared by an Egyptian Association gathering all stakeholders who are interested in evaluation in Egypt. This should never be done on an organizational level because there will be discrepancies between the organizations which will lead to the same ongoing dilemma. Organizations shall approve the cross-dictionary prepared and require future evaluations to work with it. It was noticed that each organization has its own dictionary and terminologies; However, it is not helping when Meta evaluations or Thematic evaluations are conducted. The recommendation can be implemented by EvalMENA and adopted by the Egyptian office, to ensure being dependent from any international organization.

### 7.2.2: Cross-organizational Learning

It is imperative to acknowledge that project designs are not exclusive and have been adopted by multiple organizations, both inside and outside the same entity. In order to provide practitioners with insights into previous experiences and effective interventions, evaluators have a role to play in knowledge transfer. The information gathered highlights how organizations' project activities, outputs, and outcomes are similar to one another. Therefore, promoting

knowledge sharing and the sharing of evaluation results both inside an organization and with other development partners does not imply that ideas are at risk of being stolen; rather, it highlights the significant opportunity for growth and learning. This can happen through scheduled meetings and workshops for evaluators in Egypt led by one association for all evaluations in the country to support youth and beginners navigate the field and experienced evaluators to play a role in mentoring the youth. Online and offline learning mechanisms would be essential as well as organizations to promote these cross-organizational learning workshops and events. Institutionalizing evaluation is a crucial step to happen by an independent center to be fair and balance between the different organizations.

### 7.2.3: Strengthening Ownership and Independence

Government entities shall utilize this open channel of communication between organizations to standardize the indicators that can be counted for the national standard indicators and learning outcomes. This will increase country ownership and verify that these projects are aligned with the larger local and national strategies for the country and that they are not only tied up with development partners' objectives.

OECD Aid-effectiveness working group promote for country ownership and the goal to support countries being more independent; thus, the interviewees from PEMA recommended an independent approach while conducting evaluations where evaluators are able to report freely to public institutions affiliated with the interventions as well international organizations funded the project in an attempt to increase country ownership towards achieving developmental goals.

### 7.2.4: Developing Capabilities

*M&E Orientation for All Staff Members:* The study suggests providing thorough M&E training to all staff members, including project managers, development practitioners, and members of the implementation team. The foundations of evaluations, the procedures involved, and the expectations of the commissioning team and independent evaluators should all be covered in this training. It is important to recognise the limitations of overly detailed evaluation questions, highlighting the necessity of clear, targeted questions. Important elements also include making sure report quality meets donor expectations, adequately preparing prior to the evaluator's arrival, and benchmarking against established standards.

*Project Managers:* The study highlights the significance of project managers' comprehension of M&E, recognising their critical role in the process. The integration of evaluation into project design and planning is ensured when project managers are equipped with this knowledge. Project managers have the final say in decisions, which emphasizes how important it is that they understand and value M&E processes.

*Partners in Public Institutions and Implementation:* The study emphasizes the necessity of giving capacity-building initiatives more focus, particularly for public partners. There is a critical gap in public institutions capacity because there haven't been any significant initiatives since the NGO and partners Support Centre Project (2005) (Interviewee Four, 2023). Closing this knowledge gap is crucial to guaranteeing a thorough comprehension of evaluation components, such as the purpose (why), goals (what), and approach (how). The study highlights the unsatisfactory lack of interest shown by public partners in the significance of evaluation, highlighting the need for capacity-building initiatives catered to their particular requirements.

*Employees in M&E:* One of the suggestions made for M&E employees is to start taking specific courses to improve their ability to conduct assessments. The research highlights the dearth of choices for those looking for advanced and quantitative methods courses, underscoring the need for more educational opportunities in this area.

*Independent Evaluators:* The study points out difficulties with the evaluation procedure, especially with reference to sampling techniques. One of the main recommendations is to increase the ability of evaluators to deal with problems like stratified or random sampling. Furthermore, the research highlights the significance of cultivating a cooperative atmosphere, motivating establishments to confront assessors and participate in significant dialogues concerning the extent and style of analysis.

### 7.2.5: Participatory and Inclusive Evaluations

Redefining and expanding the range of stakeholder participation in the assessment process is essential, going beyond outside determinations of indicators. Recognise the value of using participatory techniques and local viewpoints when creating assessment indicators. This methodology leverages the community's own insights to enable a more inclusive and accurate assessment of empowerment. By putting such participatory practices into practice, local groups are empowered to evaluate themselves and the evaluation process is guaranteed to be inclusive of the various viewpoints and experiences of the stakeholders.

# References

Abdelhamid, R. N. (2005). Monitoring and evaluation in Egypt. *The American University in Cairo: Master's Thesis. AUC Knowledge Fountain.* https://fount.aucegypt.edu/retro_etds/1846?utm_source=fount.aucegypt.edu%2Fretro_etds%2F1 846&utm_medium=PDF&utm_campaign=PDFCoverPages

Adams, A. (2010). *The Mubarak Kohl Initiative-Dual System in Egypt: an assessment of its impact on the school to work transition.* VOCEDplus. https://www.voced.edu.au/content/ngv%3A58895

AmCham Egypt. (2021). *US Foreign Assistance To Egypt.* AmCham Egypt. https://www.amcham.org.eg/information-resources/trade-resources/egypt-us-relations/us-foreign-assistance-to-egypt

American Evaluation Association. (2001). *What is Evaluation?* American Evaluation Association. https://www.eval.org/Portals/0/What%20is%20evaluation%20Document.pdf

Astbury, B., & Leeuw, F. L. (2010). Unpacking Black Boxes: Mechanisms and Theory Building in Evaluation. *American Journal of Evaluation*, *31*(3), 363-381.

Baker, J. L. (2000). *Evaluating the impact of development projects on poverty A handbook for practitioners.* The World Bank Washington, D.C.

Behn, R. D. (2003). Why Measure Performance? Different Purposes Require Different Measures. *Public Administration Review*, *63*(5), 586-606. 10.1111/1540-6210.00322

Bertrand, A., Mercy Corps, Beauvy-Sany, M., Education Development Center, Cilimkovic, S., Partner Microcredit Foundation, Conklin, S., Save the Children, & Jahic, S. (2009). *Monitoring and Evaluation for Youth Workforce Development Projects* [Technical Note]. The SEEP Network.

Coffman, J. (2004). *Reflecting on the Past and Future of Evaluation.* Harvard Family Research Project. https://archive.globalfrp.org/evaluation/the-evaluation-exchange/issue-archive/reflecting-on-the-past-and-future-of-evaluation/michael-scriven-on-the-differences-between-evaluation-and-social-science-research

Colwell, R. (1998). A long trail a winding: Issues, interests, and priorities in arts education research. *Arts Education Policy Review*, *99*(5), 21-30.

Crafts, N. F.R. (2011). The Marshall Plan : a reality check. *University of Warwick*, (49), 16.

Dang, H.-A. H., & Serajuddin, U. (2020). Tracking the sustainable development goals: Emerging measurement challenges and further reflections. *World Development*, *127*. 10.1016/j.worlddev.2019.05.024

Dhaliwal, I. (2022). *About Us | The Abdul Latif Jameel Poverty Action Lab.* Poverty Action Lab. https://www.povertyactionlab.org/about-us

Dichter, T. W. (2002). *Despite Good Intentions: Why Development Assistance to the Third World Has Failed.* University of Massachusetts Press.

Duflo, E., & Banerjee, A. V. (2011). *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty.* PublicAffairs.

*Egypt Impact Lab: An embedded lab in the Ministry of Planning and Economic Development.* (2022). Poverty Action Lab. https://www.povertyactionlab.org/egypt-impact-lab?lang=es

*Egypt Literate Village, Early Grade Reading Assessment (EGRA) Baseline Report.* (2022). Save the Children's Resource Centre.

https://resourcecentre.savethechildren.net/document/egypt-literate-village-early-grade-reading-assessment-egra-baseline-report/

Elbaradei, L. (2007). New Trends in Management of Official Development Assistance. *The Arab Organization for Administrative Development*. http://search.mandumah.com/Record/122895

El Baradei, L., Abdelhamid, L., & Wally, N. (2014). Institutionalising and streamlining development monitoring and evaluation in post-revolutionary Egypt: A readiness primer. *African Evaluation Journal*, *2*(1), 16. http:// dx.doi.org/10.4102/aej. v2i1.57

EvalCommunity. (2023). *Why evaluation is needed*. EvalCommunity. Retrieved April 15, 2023, from https://www.evalcommunity.com/career-center/why-evaluation-is-needed/?utm_source=sendinblue&utm_campaign=New%20editor%201800%20opportunities%20in%20ME%20and%20Impact%20Assessment&utm_medium=email

EvalMENA. (2024). *EvalMena Conference 2024: Institutionalization of Evaluation in Arab Countries*. Eval Forward. https://www.evalforward.org/events/evalmena-2023

*Evaluation Policy | Strategy and Policy | U.S. Agency for International Development*. (2020, October 1). USAID. Retrieved January 14, 2024, from https://www.usaid.gov/policy/evaluation

Eyben, R. (2013). Building Relationships in Development Cooperation: Traditional Donors and the Rising Powers. *IDS Policy Briefings*, (36).

GIZ. (2021). Central Project Evaluation. *MADIBA Consult GmbH*. https://d-nb.info/123313647X/34

GIZ. (2023). *Egypt - giz.de*. GIZ. https://www.giz.de/en/worldwide/319.html

Glewwe, P., & Todd, P. (2022). *Impact Evaluation in International Development*. World Bank Group.

Haines, J. (2023, October 13). *Countries That Receive the Most Foreign Aid From the U.S. | Best Countries | U.S. News*. U.S. News & World Report. Retrieved December 8, 2023, from https://www.usnews.com/news/best-countries/articles/countries-that-receive-the-most-foreign-aid-from-the-u-s

Hoefer, R. (2000). Accountability in Action?: Program Evaluation in Nonprofit Human Service Agencies. *Nonprofit Management and Leadership*, *11*(2), 167-177. https://onlinelibrary.wiley.com/doi/10.1002/nml.11203

Hogan, R. L. (2007). The Historical Development of Program Evaluation: Exploring Past and Present. *Online Journal for Workforce Education and Development*, *2*(4).

Holzapfel, S. (2016). Boosting or Hindering Aid Effectiveness? An Assessment of Systems for Measuring Donor Agency Results. *Public Administration and Development*, *36*(1), 3-19. 10.1002/pad.1749

House, E. R. (1994). The Future Perfect of Evaluation.

Kabonga, I. (2018). Principles and Practice of Monitoring and Evaluation: A Paraphernalia for Effective Development. *Journal of Development Studies*. https://doi.org/10.25159/0304-615X/3086

Kanyamuna, V., & Phiri, M. (2019). Who said Monitoring and Evaluation is not rooted in firm theoretical foundations? A review of relevant literature. *International Journal of Humanities, Art and Social Studies*, *1*(4), 1-22. 10.12691/wjssh-6-1-4

Kariuki, J. (2014). An Exploration of the Guiding Principles and Challenges of Monitoring and Evaluation of Community Development Projects and Programmes. *International Journal of Business and Social Science*, *5*(1), 140-147.

Karlan, D., & Appel, J. (2012). *More Than Good Intentions: Improving the Ways the World's Poor Borrow, Save, Farm, Learn, and Stay Healthy*. Penguin Publishing Group.

Koolwal, G. B., Samad, H. A., & Khandker, S. R. (2010). *Handbook on Impact Evaluation: Quantitative Methods and Practices*. World Bank.

KPMG International. (2014). Monitoring and Evaluation in the Development Sector.

Kremer, M., Gallant, S., Rostapshova, O., & Thomas, M. (2021). Is Development Economics a Good Investment? Evidence on scaling rate and social returns from USAID's innovation fund. *Harvard University*. https://cpb-us-w2.wpmucdn.com/voices.uchicago.edu/dist/0/2830/files/2021/09/SROR-21.03.12_clean-3.pdf

Lele, U. (1990). Agricultural Growth and Assistance to Africa: Lessons of a Quarter Century. *San Francisco: International Center for Economic Growth*, 7-9.

Liket, K. C., Rey-Garcia, M., & H. Maas, K. E. (2014). Why Aren't Evaluations Working and What to Do About It: A Framework for Negotiating Meaningful Evaluation in Nonprofits. *American Journal of Evaluation*, *35*(2), 171-188. http://journals.sagepub.com/doi/10.1177/1098214013517736

Linfield, K. J., & Posavac, E. J. (2018). *Program Evaluation: Methods and Case Studies*. Routledge.

Locke, E. A. (1982). The Ideas of Frederick W. Taylor: An Evaluation. *The Academy of Management Review*, *7*(1), 14-24.

Lune, H., & Berg, B. (2016). *Qualitative Research Methods for the Social Sciences*. Pearson.

Madaus, G. F., Kellaghan, T., & Stufflebeam, D.L. (Eds.). (2005). *Evaluation Models: Viewpoints on Educational and Human Services Evaluation*. Springer Netherlands.

Marshall, C., & Rossman, G. B. (2014). *Designing Qualitative Research*. SAGE Publications.

Marshall, J. H., & Suárez, D. (2014). The Flow of Management Practices: An Analysis of NGO Monitoring and Evaluation Dynamics. *Nonprofit and Voluntary Sector Quarterly*, *43*(6).

McCloskey, S. (2015). From Mdgs to Sdgs: We Need a Critical Awakening to Succeed. *Policy & Practice: A Development Education Review*.

Ministry of Planning and Economic Development. (2021). *EGYPT'S 2021 VOLUNTARY NATIONAL REVIEW*. Sustainable Development Goals. https://sustainabledevelopment.un.org/content/documents/279512021_VNR_Report_Egypt.pdf

Mohamed, R. K. (2023). The Dual Education System in Egypt: the system's challenges and the way forward. *AUC Knowledge Fountain*. https://fount.aucegypt.edu/etds/2069

Mulligan, M., & Sherriff, L. (2019). *MEAL DPro*. Humanitarian Leadership Academy. https://www.scribd.com/document/589998910/MEAL-DPro-Guide-Eng

Natsios, A. (2010). The clash of the counter-bureaucracy and development. *Center for Global Development*. www.cgdev.org/content/publications/detail/1424271

OECD. (1991). *Evaluation Criteria*. OECD. https://www.oecd.org/dac/evaluation/daccriteriaforevaluatingdevelopmentassistance.htm

OECD. (2007). *Aid Effectiveness 2006 Survey on Monitoring the Paris Declaration Overview of the Results*. OECD.

OECD. (2011). Egypt Chapter. In *Aid Effectiveness: Progress in Implementing The Paris Declaration*.

OECD. (2019). *The High Level Fora on Aid Effectiveness: A history*. OECD. https://www.oecd.org/dac/effectiveness/thehighlevelforaonaideffectivenessahistory.htm

OECD. (2023, March). *EGYPT-OECD Country Programme* [Progress Report from September 2022 – March 2023]. OECD.

Orr, L. L. (2018). The Role of Evaluation in Building Evidence-Based Policy. *American Academy of Political and Social Science*, *678*, 51-59. https://www.jstor.org/stable/10.2307/26582338

Patton, M. Q. (1978). *Utilization-focused evaluation*. SAGE Publications.

Ravallion, M. (2009). Evaluation in the Practice of Development. *Oxford University Press*, *24*(1), 29-53. https://www.jstor.org/stable/40282377

Rist, R. C., & Morra Imas, L. G. (2009). *The Road to Results: Designing and Conducting Effective Development Evaluations*. World Bank. 10.1596/978-0-8213-7891-5

Roob, N. (2014). *The Value of Rigorous Evaluation*. Stanford Social Innovation Review. https://ssir.org/articles/entry/the_value_of_rigorous_evaluation

Sachs, J. (2006). *The End of Poverty: Economic Possibilities for Our Time*. Penguin Publishing Group.

*Sample Size Calculator*. (2023). Qualtrics. https://www.qualtrics.com/blog/calculating-sample-size/

Scriven, M. (1981). *Evaluation Thesaurus*. ERIC.

Scriven, M. (1996). The theory behind practical evaluation. Evaluation. *2*(4), 393-404.

Sette, C. (2021, November 3). *Participatory evaluation*. Better Evaluation. Retrieved November 20, 2023, from https://www.betterevaluation.org/methods-approaches/approaches/participatory-evaluation

Stern, J., Davis, J., & Vukmirovic, Z. (2016). *EGRA Toolkit*. DEVELOPMENT EXPERIENCE CLEARINGHOUSE | PDF Server. https://pdf.usaid.gov/pdf_docs/PA00M4TN.pdf

Thomas, J. C., Doherty, K., Grant, S. W., & Kumar, M. (2021). Advances in monitoring and evaluation in low- and middle-income countries. *Evaluation and Program Planning*, *89*. https://www.sciencedirect.com/science/article/abs/pii/S0149718921000896

Toscano, S. (2013). Exploring the History and Challenges of Monitoring and Evaluation in International Nongovernmental Organizations: Complemented by Intern Experience at Save the Children USA. *SIT Graduate Institute*.

UN. (2000). *United Nations: Millenium Declaration*. the United Nations. Retrieved April 23, 2023, from https://www.un.org/en/development/devagenda/millennium.shtml

UN. (2015). Resolution adopted by the General Assembly on 25 September 2015. *UN*.

UN. (2018). *The United Nations in Egypt*. United Nations in Egypt. Retrieved December 9, 2023, from https://egypt.un.org/en/about/about-the-un

UNICEF. (2003). Programme Policy and Procedures Manual: Programme Operations.

UN Task Force. (1984). Guiding principles for the design and use of monitoring and evaluation in rural development projects and programmes. https://digitallibrary.un.org/record/1079?ln=en

USAID Egypt. (2020). *Country Development Cooperation Strategy (CDCS)*. USAID. https://www.usaid.gov/sites/default/files/2022-05/CDCS-Egypt-December-2025.pdf

Valadez, J. J., & Bamberger, M. (1994). *Monitoring and Evaluating Social Programs in Developing Countries: A Handbook for Policymakers, Managers, and Researchers* (J. J. Valadez & M. Bamberger, Eds.). World Bank.

Vandemoortele, J. (2020). *Are the SDGs a major reboot or a sequel to the MDGs?* OECD.

Waldrop, M. M. (1992). *Complexity: the emerging science at the edge of order and chaos*. Simon & Schuster.

World Bank. (2011). Monitoring and Evaluation Capacity Development. The World Bank.

Yonehara, A., Saito, O., Hayashi, K., Nagao, M., Yanagisawa, R., & Matsuyama, K. (2017). The role of evaluation in achieving the SDGs. *Sustainability Science*, *12*(6), 969-973. https://doi.org/10.1007/s11625-017-0479-4

Zahran, N. (2023, June). Assessing The Implementation of Egypt's Education 2.0 Reform: Assessing The Implementation of Egypt's Education 2.0 Reform: The Case of Sohag. *The American University in Cairo: Master's Thesis. AUC Knowledge Fountain.* https://fount.aucegypt.edu/cgi/viewcontent.cgi?article=3152&context=etds

## *Appendices*

Appendix One:

Atlas.ti coding

graph



Figure 4: *The most repeated codes and the number of times repeated along all interviews.*
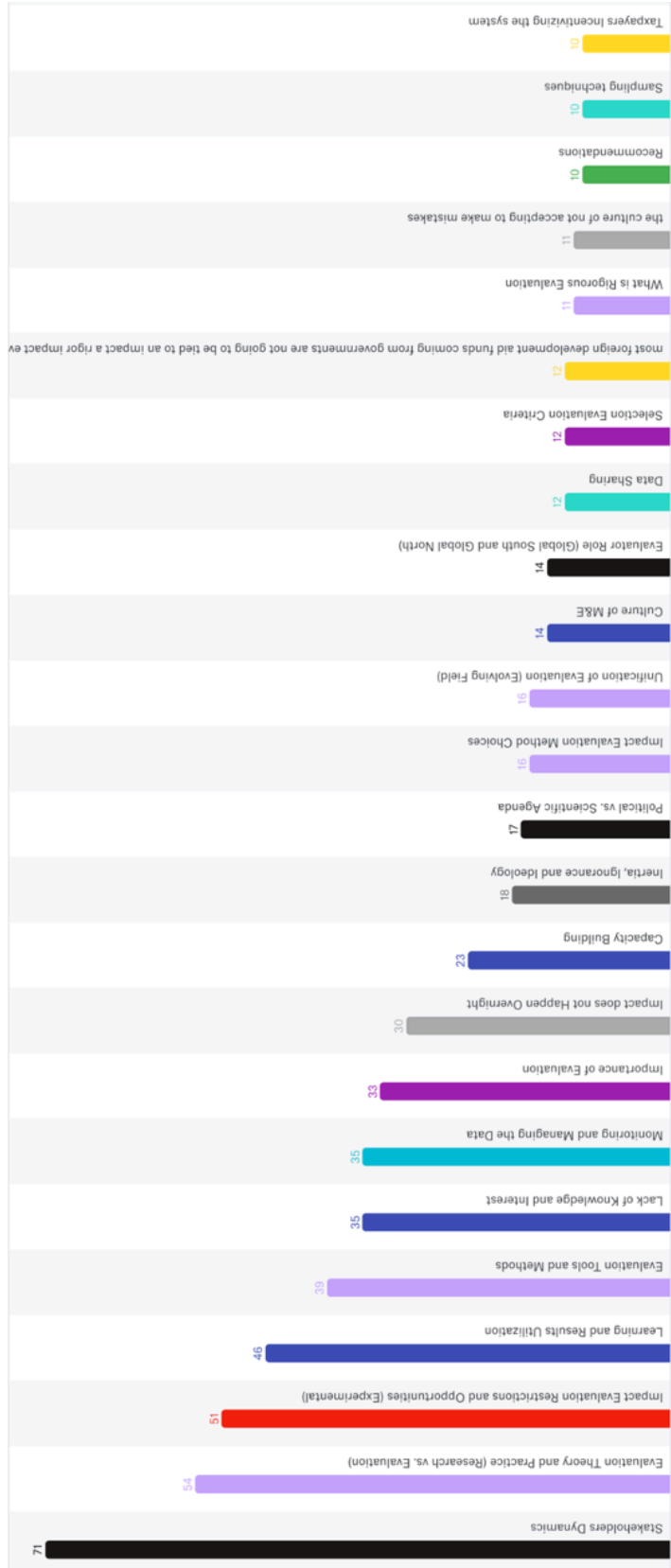
Resource: Screenshot from Atlas.ti website personal account of the researcher.

*Note: The codes were merged and classified into different order than in the figure to represent a meaningful flow of information*

**Demographics:**

- Type: (Independent Evaluator, Development Partner, Government Official)
- Interviewee title:
- Years of experience:
- Gender:
- Duration:

**Purpose of the thesis to** Identify the opportunities and challenges involved in evaluating development projects in Egypt. The study aims to explore the current evaluation approaches to measure all levels of changes whether impact, output or outcome used in Egypt, know their strengths and weaknesses, and provide recommendations for improving the evaluation practices for development partners (bilateral) to ensure better accountability, transparency, and learning. The study seeks to contribute to the literature on program evaluation in the context of developing countries, particularly in Egypt, and to inform policy and practice for effective project evaluation.

**Questions directed to an (Independent Evaluator)**

**Background information:**

*"Evaluation is a systematic process to determine merit, worth, value or significance. Programs and projects of all kinds aspire to make the world a better place. Program evaluation answers questions like: To what extent does the program achieve its goals? How can it be improved? Should it continue? Are the results worth what the program costs? Program evaluators gather and analyze data about what programs are doing and accomplishing to answer these kinds of questions." (American Evaluation Association, 2001) We mean in this research all types, design, methods and approaches of evaluations that have been used before in Egypt to evaluate development programs.*

1. Please introduce yourself and what is your experience with evaluating development projects in Egypt?

2. In which sectors have you been part of commissioning or planning for evaluations in the organization? And which government entities/ development partners?

3. What do you think are the differences between evaluations in academia and research and in practice in development organizations?

**Theme 1: Planning Evaluations:**

4. For the projects you worked on, **at what stage** of the project cycle (planning, implementation...etc.) were the evaluation plans initially set?
   - Ideally, at which stage do you think evaluation plans should be set? Why?

5. **Why** do you think development partners undergo evaluations of their programs?
   - What other purposes do you think might be important yet commonly neglected by development partners?

6. Based on your experience, **who** is usually responsible for deciding on the evaluation type, questions, designs and tools for a program?
   - Ideally, who should be responsible? Why?
   - What **other factors** (objectives, budget,...) contribute to deciding on the evaluation type and design?
   - What is the **most widely evaluation** type**,** approach and design used within the organization? Why?
   - How accepting were project partners of your suggested amendments as an external evaluator to the design, techniques and data collection tools used?
   - To what extent are the public partners involved in evaluations?

7. Could you share (including challenges and limitations) a time when you were **part of the evaluation** process from the planning stage of a project, if any?

8. In what ways do you see results of evaluations get **utilized**? How do you think they should be utilized? Why?

**Theme 2: Evaluation Tools:**

9. In the evaluations you worked on, what are the **different tools** commonly used in projects funded/implemented by the organization (e.g., FGDs, Interviews, Surveys, Observation methods, others)?

10. Were there any guidance/guidelines offered by the development partner organizations or implementing partners, if any, that you needed to follow while developing these tools? For instance, specific questions/themes that you needed to include in surveys, interviews...etc.
    - Do the projects already have a Logical framework and Theory of change or what else?

**Theme 3: Evaluation Methods (Opportunities and Challenges)**

11. Are you familiar with the differences between **experimental and non-experimental** evaluations?
    - Have you worked on experimental and/or non-experimental evaluations before?

12. Have you conducted **experimental evaluations** yourself or heard of ones that were conducted before? If so, can you briefly describe the evaluation(s)? Can you cite examples?
    - If we think of experimental evaluations, what are the opportunities?
    - And what are the challenges?

13. Have you conducted **quasi-experimental** yourself or heard of ones that were conducted evaluations before? If so, can you briefly describe the evaluation(s)? Can you cite examples?
    - In terms of quasi-experimental, what are the opportunities?
    - And what are the challenges?

14. Have you conducted **non-experimental evaluations** (Mixed, Qual, Quan) before? If so, can you briefly describe the evaluation(s)? Can you cite examples?
    - And for the non-experimental, what are the opportunities?

- And what are the challenges?

15. Have you conducted any evaluations that targeted **evaluating the impact (DAC Criteria)** of a project/program? If so, can you briefly describe the evaluation(s)? Can you cite examples?

    - What are the opportunities?

    - And what are the challenges?

    - How should organizations following DAC criteria for evaluation should measure impact and sustainability?

**Theme 4: Rigorous Vs. Non-rigorous**

According to literature, "the essential characteristic of a rigorous evaluation is that it includes a credible strategy for identifying the counterfactual to measure the impact of the project." (Ravallion, 2009, p. 31).

16. As an evaluator, what are your thoughts regarding this quote?

    - How do you **define rigorous and non-rigorou**s evaluations?

17. What were the **challenges encountered** or heard of if you happened to design or learn about a rigorous evaluation? based on your experience within the Egyptian context and the development projects implemented in Egypt? What were the opportunities?

**Final remarks**

18. What are your **recommendations for improving** the process of Evaluation in development partners projects to **achieve effective development** in Egypt?

19. In your opinion, do you see differences while practicing evaluation between donors, governmental entities and sectors?

    - To what extent are the public partners involved in evaluations?

20. Is there anything else you would like to comment on regarding 'Evaluation' as carried out by development partners in partnership with public entities?

21. Would you like to recommend someone in academia/practice to be invited to the study?

**Questions directed to a** Development partner/ Government Official

**Background information:**

**What is Evaluation from literature:**

> *"Evaluation is a systematic process to determine merit, worth, value or significance. Programs and projects of all kinds aspire to make the world a better place. Program evaluation answers questions like: To what extent does the program achieve its goals? How can it be improved? Should it continue? Are the results worth what the program costs? Program evaluators gather and analyze data about what programs are doing and accomplishing to answer these kinds of questions." (American Evaluation Association, 2001) We mean in this research all types, design, methods and approaches of evaluations that have been used before in Egypt to evaluate development programs.*

1. Please introduce yourself and what is your experience with evaluating development projects in Egypt?

2. In which sectors  have you been part of commissioning or planning for evaluations in the organization? And which government entities/ development partners?

3. What do you think are the differences between evaluations in academia and research and in practice in development organizations?

4. What was the relationship between public organizations and development partners? (you have worked with)
   - Who reports to whom?
   - If a project was commissioned by an international development partner and partnered with a ministry, who does the M&E?

**Theme 1: Planning Evaluations:**

5. For the projects you worked on, **at what stage** of the project cycle (planning, implementation...etc.) were the evaluation plans initially set?
   - Ideally, at which stage do you think evaluation plans should be set? Why?

6. **Why** do you think development partners undergo evaluations of their programs?
    - What other purposes do you think might be important yet commonly neglected by development partners?
7. Based on your experience, **who** is usually responsible for deciding on the evaluation type, questions, designs and tools for a program?
    - Ideally, who should be responsible? Why?
    - What **other factors** (objectives, budget,...) contribute to deciding on the evaluation type and design?
    - What is the **most widely evaluation** type**,** approach and design used within the organization? Why?
8. Could you share (including challenges and limitations) a time when you were **part of the evaluation** process from the planning stage of a project, if any?
9. In what ways do you see results of evaluations get **utilized**? How do you think they should be utilized? Why?

**Theme 2: Evaluation Tools:**

10. In the evaluations you worked on, what are the **different tools** commonly used in projects funded/implemented by the organization (e.g., FGDs, Interviews, Surveys, Observation methods, others)?
11. Were there any **guidance/guidelines** offered by your organization or implementing partners, if any, that evaluators needed to follow while developing these tools? For instance, specific questions/themes that you needed to include in surveys, interviews...etc.
    - Do the projects already have a Logical framework and Theory of change or what else?

**Theme 3: Evaluation Methods (Opportunities and Challenges)**

12. Are you familiar with the differences between **experimental and non-experimental** evaluations?
    - Have you worked on experimental and/or non-experimental evaluations before?

13. Has your organization decided on conducting **experimental evaluations** before ? If so, can you briefly describe the evaluation? Can you cite examples?
    - What were the opportunities?
    - What were the challenges?

14. Has your organization decided on conducting **quasi-experimental** evaluations before? If so, can you briefly describe the evaluation? Can you cite examples?
    - What were the opportunities?
    - What were the challenges?

15. Has your organization decided on conducting **non-experimental** evaluations before? If so, can you briefly describe the evaluation? Can you cite examples?
    - What were the opportunities?
    - What were the challenges?

16. Has your organization decided on conducting any evaluations that targeted **evaluating the impact** of a project/program? If so, can you briefly describe the evaluation(s)? Can you cite examples?
    - How should organizations following DAC criteria for evaluation should measure impact and sustainability?
    - Are there any guidelines about that from DAC?

**Theme 4: Rigorous Vs. Non-rigorous**

According to literature, "the essential characteristic of a rigorous evaluation is that it includes a credible strategy for identifying the counterfactual to measure the impact of the project." (Ravallion, 2009, p. 31).

17. As someone affiliated with an organization in practice, what are your thoughts regarding this quote?
    - How do you define/ perceive rigorous and non-rigorous evaluations?

18. What were the **challenges encountered** or heard of if you happened to design or learn about a rigorous evaluation? based on your experience within the Egyptian context and the development projects implemented in Egypt.

- What were the opportunities?

**Final remarks**

19. What are your **recommendations for improving** the process of Evaluation in development partners projects to **achieve effective development** in Egypt?
20. Would you like to recommend someone to be invited to the study?