Building M&E Capacity for SFSD

Promising Practices, Assessment of Current System, and Recommendations for the Future

مؤسسة ساويرس للتنمية الاجتماعية

Sawiris Foundation For Social Development

University of Minnesota

Humphrey School of Public Affairs
University of Minnesota

Interdisciplinary Center for the Study of Global Change

FINAL REPORT

Humphrey School Team

Felipe Dyna Barroso – Master of Development Practice Candidate, 2016

Trupti Sarode – Master of Development Practice Candidate, 2016

Yue Yu – Master of Public Policy Candidate, 2017

Faculty Advisers

Dr. Robert Kurdle – Humphrey School of Public Affairs, Freeman Center for International Economic Policy

Dr. Sherry Gray – Humphrey School of Public Affairs, Global Policy Area

Supporting Organizations

Sawiris Foundation for Social Development

University of Minnesota | Humphrey School of Public Affairs

Interdisciplinary Center for the Study of Global Change (ICGC)

Report Prepared By: Humphrey School Team

Building M&E Capacity

Table of Contents

Building M&E Capacity

Glossary

Organization Acronyms

3ie – International Initiative for Impact Evaluation
DOL – U.S Department of Labor
EFE – Education for Employment
HHH – Humphrey School of Public Affairs
IYF – International Youth Foundation
J-PAL – Abdul Latif Jameel Poverty Action Lab
SFSD – Sawiris Foundation for Social Development
YEI – Youth Employment Inventory

Monitoring and Evaluation Terms

Cause-and-Effect Evaluations – Evaluations that attempts to segregate the impact of external factors from the impact of the program.
CBA – Cost-Benefit Analysis
CEA – Cost-Efficiency Analysis
Counterfactual – Hypothetical scenario of what would have happened had the program not taken place.
Descriptive Evaluations – Evaluations that tell us what is taking place in the program.
Experimental Design – An impact evaluation that randomizes the selection of beneficiaries.
External Validity – The extent to which the results of the evaluation can be generalized.
Falsification Test – Test that accounts for potential differences between the trends in the outcomes of control and treatment. It requires a pre-baseline survey.
Internal Validity – The extent to which the evaluation correctly answers the questions it claims to answer regarding what is being evaluated.
Logic Model – Visual depiction of how the program is supposed to work, the theory of the program, and the relationship between resources, activities, and outcomes (short- and long-term).
M&E – Monitoring and Evaluation
MENA – Middle-East and North Africa region
Normative Evaluations – Evaluations that tell us what is taking place, and what should be taking place.
Performance Measurement System - Performance measurement system is a set of tools, practices, and processes that collects, analyzes, and reports key information regarding the performance of an organization. Indicators that are used in several individual program evaluations are often made part of the performance measurement system
QOE – Quality of Evaluation
SCBA – Social Cost-Benefit Analysis
Selection Bias – Non-random selection of participants for the program, potentially resulting in systematic differences between the control and treatment groups.
Theory of Change – Explicit theory of how a program contributes to a set of specific outcomes through a series of intermediate results.

Building M&E Capacity

Executive Summary

This report, titled "*Building M&E Capacity for SFSD: Promising Practices, Assessment of Current System, and Recommendations for the Future*", is the result of a six month-long consulting project provided by three graduate students from the University of Minnesota (Humphrey School Team) to Sawiris Foundation for Social Development (SFSD).

The goal of this report is to assist SFSD in its efforts to build the necessary monitoring and evaluation (M&E) capacity as a means to strengthen its portfolio of programs. For concision purposes, this report focuses on SFSD's youth employment training programs, but many of the M&E principles presented here are applicable on different programmatic areas as well.

Six strategic recommendations are put forth at the end of the report (Part VI). They are:

1. Create an M&E Position in SFSD

   Such employee would need to be able to: a) decide, from a technical perspective, which projects need to be evaluated, b) decide which type of evaluation to use on the selected projects, c) conduct the implementation of simple evaluation projects, such as process or outcome evaluations, d) help operate the performance measurement system of SFSD and, e) be the point-of-contact with external organizations & consultants in issues that relate to M&E.

2. Establish a Performance Measurement System

   While SFSD currently keeps track of some information (activities & outcomes) of the programs it funds, we recommend that the foundation creates a centralized database to curate key indicators from most or all of its programs—thus constituting a performance measurement system.

3. Use a Humphrey School of Public Affairs "Field Experience Team" to Carry Out a Process/Outcome Evaluation of the Vocational Productivity Program.

   The Humphrey School Team recommends that SFSD implements the draft evaluation plan for the *Vocational Productivity Program* in collaboration with a Field Experience Team from the Humphrey School of Public Affairs. Such teams are composed by graduate students and come at very low cost to SFSD.

4. Partner With Specialized Organizations When an Impact Evaluation is Needed.

   Conducting an impact evaluation requires specialized skills and knowledge in quantitative data collection and analysis. For this reason, the Humphrey School Team recommends that SFSD partners with reputed organizations such as the Taqeem Initiative, J-PAL, or the International Initiative for Impact Evaluation when the need for an impact evaluation is identified.

5. Integrate Program Evaluation Into Program Design.

   Embedding M&E practices into the program design of the programs funded by SFSD is a crucial factor for reducing the cost of evaluations as well as for strengthening their findings. Evaluations planned for in advance have no downside when compared to retrospective evaluations.

6. Share the Knowledge When Robust Evaluations are Carried Out

There is a significant gap of evidence when it comes to youth employment training programs in MENA. If SFSD intends to strengthen its M&E practices, the Humphrey School Team recommends that SFSD shares the findings of robust evaluation projects among other organizations as a means to further contribute to its mission.

Besides the recommendations covered in Part VI, this report serves as a cursory guide to M&E practices and methods. The purpose is to provide SFSD's staff with enough information in order to be able to make informed strategic decisions regarding the creation of M&E capacity. Parts I through V of the report are a mixture of explanations of M&E concepts with their application to SFSD's context.

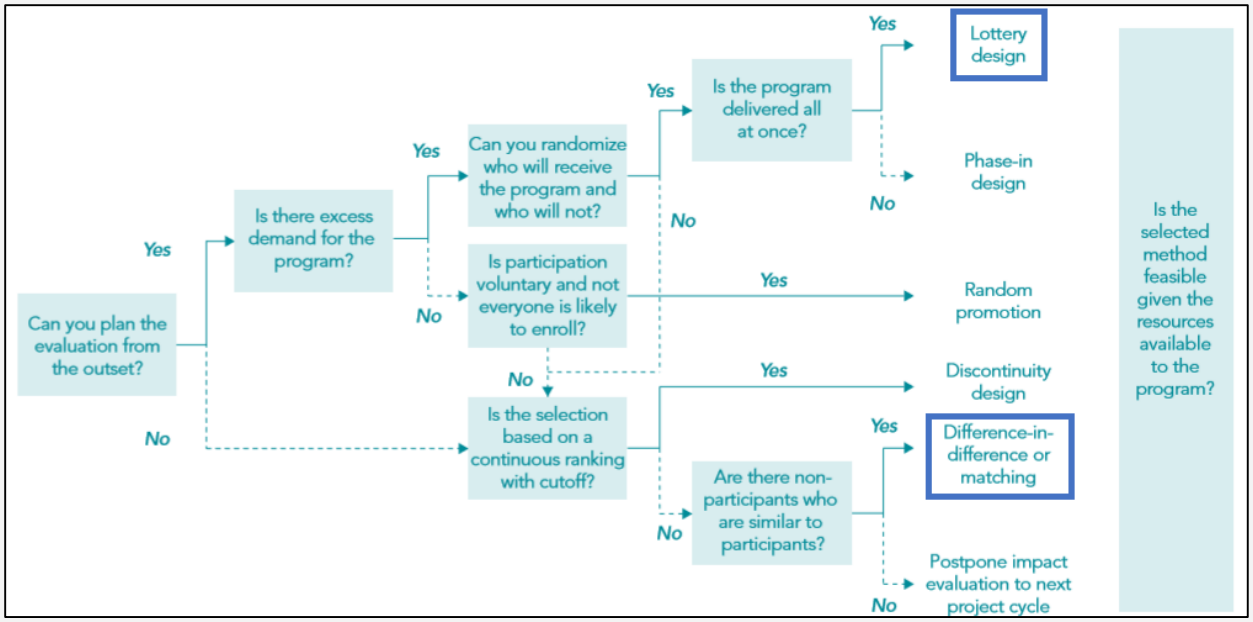
Part I of the report provides an overview of the types of program evaluation that could be used by SFSD in order to evaluate different aspects of a particular program. While the M&E literature may have additional types of program evaluation not considered in this report, the Humphrey School Team selected five types that were considered crucial for SFSD's context. They are: a) monitoring, b) process evaluation, c) outcome evaluation, d) impact evaluation, and e) cost-effectiveness/benefit analysis or social cost-benefit analysis. The table below summarizes the information presented in the overview:

Evaluation Types: Summary Table

Evaluation Type	What does it tell us?	Organizational Capacity	Estimated Costs*	Estimated Duration
Monitoring	Descriptive information of the program (such as participant's engagement, etc.)	Low	Unspecified	Unspecified
Process Evaluation	Whether the program is being implemented as intended or not (and why)	Moderate	$10,000 - $60,000	1-6 months
Outcome Evaluation	The change in short-term outcomes of the program (compares w/ pre-established goals)	Moderate	$10,000 - $30,000	1-3 months
CEA/CBA/SCBA	The outcome (non-financial, financial, social) per unit of cost ratio	Moderate	+ $10,000 - $30,000	+ 1-3 months
Impact Evaluation	Impact of the program on medium- or long-term outcomes, separated from external factors	High	$15,000 - $1+ million	6 months (retrospective) 1-2 years (prospective)
CEA/CBA/SCBA	The impact (non-financial, financial, social) per unit of cost ratio	Moderate	+ $10,000 - $30,000	+ 1-3 months

**for a detailed explanation of estimated costs, refer to Part II*

Part II of the report details each type of program evaluation introduced in Part I, covering the objectives, the estimated duration, the financial/human/technology resources required, and providing an example of each type of evaluation for illustration purposes. Special attention is given to Impact Evaluation through an explanation of the methods that allow evaluators segregate the impacts of external factors from the impacts of the program. Below is a decision tree to help SFSD's staff decide which Impact Evaluation method to use, which is also included at the end of the section.



Adapted from The World Bank's "Measuring Success of Youth Livelihood Interventions (2012) report

Part III of the report analyzes the M&E practices of three case studies from analogous organizations as means to stimulate SFSD staff's thinking in terms of what indicators and what data collection methods to use when designing its own program evaluations. Part IV complements Part III by situating the current M&E practices of SFSD in the analytical framework of this report as well as juxtaposing them with the practices of the selected case studies.

Part V presents a draft evaluation plan for the *Vocational Productivity Program* while documenting the process that creating such a plan entails. The objective is to help SFSD's staff to understand the steps that are necessary when planning for the evaluation of a program. The draft plan includes a Logic Model of the *Vocational Productivity Program*, a stakeholder analysis, the key evaluation questions, and suggested indicators and data collection methods. It must be noted, however, that the evaluation plan was not designed in-locus and, therefore, needs to be reviewed by the local M&E Manager before implementation.

Lastly, **Appendix A** lists a series of resources regarding youth employment programs and M&E practices that may be of use for SFSD's staff (such as workshops, training courses, evaluation databases, consultant databases, etc.)

Part I

Overview of M&E Options

Building M&E Capacity

Part I: Overview of M&E Options

Purpose of Part I

This section intends to inform SFSD's staff of the existing **types of program evaluation** that are applicable to SFSD's portfolio. The document covers the purposes, resource requirements (financial or not), expected duration, and limitations of each type of program evaluation. Each type is also accompanied by a real-world example for illustration purposes.

Part I is a summary of Part II, which describes each type of evaluation in detail. The framework of analysis that is the basis of both parts is an adaptation of the framework used in the World Bank's report titled "*Measuring Success of Youth Livelihood Interventions*" (Hempel & Fiala, 2011). Adaptations were made to better fit SFSD's context as well as other sources that were used in this report.

M&E Options' Summary

Descriptive Evaluations: tell us what's going on in our program

Evaluation Type 1: Monitoring

Purpose: Monitoring is the simplest type of evaluation since it is purely descriptive. Monitoring consists in collecting specific information about a program through observation and reporting it at a later stage. Which information to collect is defined by the program's *theory of change*[1].

Resource Requirements:

Estimated Cost[2] – 5% of program's overall costs.
Main Cost Drivers – Salary of M&E staff; scope of evaluation.
Evaluator's Necessary Skills – Program design and monitoring.
Technology – Data management system

Expected Duration: Unspecified[3]

Limitations: Does not provide insight on the *merits* of the program.

Normative Evaluations: tell us what is taking place and what should be taking place

Evaluation Type 2: Process Evaluation

[1] Program theory is an explicit theory of how a program contributes to a set of specific outcomes through a series of intermediate results.
[2] Whenever a lump sum estimate is not available, SFSD may allocate 5% of the program's overall costs for M&E purposes. The "5% rule" is widely used by many organizations around the world.
[3] The report "Measuring Success of Youth Livelihood Interventions" did not specify the duration of monitoring evaluations.

Purpose: Process evaluations document whether a program has been implemented as intended and the reasons for correct/incorrect implementation. It focuses on the program's activities instead of its outcomes.

Resource Requirements:

Estimated Cost – $10,000 - $60,000 or 5% of program's overall costs.
Main Cost Drivers – Salary of M&E staff; scope of evaluation.
Evaluator's Necessary Skills – Program design and monitoring; qualitative data collection; qualitative data analysis.
Technology – Survey collection tools; interviewing tools; data management system.

Expected Duration: 1 – 6 months

Limitations: Does not evaluate impact or outcomes of the program.

Illustration Case: *United States Department of Labor—Youth Offender Demonstration Program (1999)*

Evaluation Type 3: Outcome Evaluation

Purpose: Outcome evaluations measure the change in short-term variables of interest (e.g. number of program participants employed, skills learned, etc.) before and after the implementation of the program and compares them with pre-established goals.

Resources Required:

Estimated Cost – $10,000 - $30,000 or 5% of program's overall costs.
Main Cost Drivers – Salary of M&E staff; scope of evaluation.
Evaluator's Necessary Skills – Program design and monitoring; quantitative data collection.
Technology – Survey collection tools; data management system

Expected Duration: 1 – 3 months

Limitations: Short-term outcomes may not translate into long-term benefits; seldom uses control groups to account for the effects of external factors.

Illustration Case: *International Youth Foundation – Youth Empowerment Program (2007)*

Cause-and-Effect Evaluations: tells us what difference the program makes

Evaluation Type 4: Impact Evaluation

Purpose: Impact evaluations estimate the effect of the program on the medium- and long-term variables of interest. It segregates the effect of the program from external factors through the use of control groups.

Resources Required:

Estimated Cost – $15,000 - $1 million or 5% of program's overall costs.
Main Cost Drivers – Extensive, rigorous data collection; M&E staff or expert consultant's salary.
Evaluator's Skills – Quantitative data collection; quantitative data analysis (advanced statistics and econometrics).

Technology – Survey collection tools; data management system; statistical software packages.

Expected Duration: 6 months for retrospective evaluation or 12-24 months for prospective evaluations

Limitations: High human capital and financial requirements; long duration; often needs to be integrated to program design (planned for in advance).

Illustration Case: *Colombian Federal Government – Youth in Action (2005)*

Integrating Cost Effectiveness/Cost-Benefit Analyses with M&E

Evaluation efforts can have their worthiness increased when coupled with cost-effectiveness (CEA), cost-benefit (CBA), or social cost-benefit (SCBA) analyses, which integrate cost considerations to the findings of the program evaluation.

These analyses, when used for a specific program, are complementary types of evaluation and are best when not carried out on their own. For instance, the findings of an outcome evaluation or an impact evaluation (e,g., increase in income of participants) can be used as inputs for a CEA/CBA/SCBA.

Evaluation Type 5: Cost-Effectiveness/Cost-Benefit/Social Cost-Benefit Analysis

Purpose: A *cost-effectiveness analysis* assesses whether or not the program is generating enough units of outcomes (e.g. jobs created) per unit of cost (e.g. Egyptian pounds).

A *cost-benefit analysis assesses* whether or not the program is generating enough units of monetary benefits to program participants (e.g. average salary increase) per unit of cost.

A *social cost-benefit analysis* assesses whether or not the program is generating enough monetary benefits to society per unit of cost (it monetizes program's effects on external factors such as national labor market, public safety, pollution, etc.).

Resources Required:

Estimated Cost – $10,000 - $30,000 or 5% of the program's overall costs.
Main Cost Drivers – M&E Staff salary.
Evaluator's Necessary Skills – Valuation.
Technology – Data management system

Expected Duration: 1 – 3 months

Limitations: The reliability of the findings of this type of evaluation is contingent on the quality of the outcome or impact evaluation that it uses as basis.

Illustration Case: *Argentina's Ministry of Labor – Youth Training Program (1999)*

Quality of Evaluation (QOE)

The Youth Employment Inventory (YEI) is a World Bank-initiated partnership between development agencies and inter-governmental organizations that seeks to improve "effective collaboration at country, regional, and global levels (…) with the aim of advancing knowledge of effective approaches to youth employment promotion."[4] The YEI compiles evaluations of interventions of youth employment training worldwide and ranks them using a "Quality of Evaluation" index. The higher the index, the more complete the information provided by the type of evaluation. The table below summarizes the QOE measurement:

QOE Value	Description
0	Program has no evaluation information available on **outcomes or impact**.
1	Evaluation includes basic information on the **gross outcomes of the intervention** (e.g., number of participants/young people who found a job after the intervention, improvement in earnings of participants) without considering **net effects** (i.e., there is no control group).
2	Evaluation includes estimate of **net impact** on, e.g., employment and earnings in the labor market (using control groups to measure impact) but no cost-benefit analysis.
3	Evaluation includes **net impact** plus **cost-benefit analysis**.

Table retrieved from *Betcherman, Godfrey, Puerto, Rother, Stavreska (2007).*

Evaluation Types: Summary Table

Evaluation Type	What does it tell us?	Findings	Organizational Capacity	Estimated Costs*	Estimated Duration	Example	QOE
Monitoring	Descriptive information of the program (such as participant's engagement, etc.)	Descriptive	Low	Unspecified	Unspecified	-	0
Process Evaluation	Whether the program is being implemented as intended or not (and why)	Normative	Moderate	$10,000 - $60,000	1-6 months	Youth Offender Demonstration - U.S Department of Labor (1999)	0
Outcome Evaluation	The change in short-term outcomes of the program (compares w/ pre-established goals)	Normative	Moderate	$10,000 - $30,000	1-3 months	Youth Empowerment Program - IYF (2007)	1
CEA/CBA/SCBA	The outcome (non-financial, financial, social) per unit of cost ratio	-	Moderate	+ $10,000 - $30,000	+ 1-3 months	-	1
Impact Evaluation	Impact of the program on medium- or long-term outcomes, separated from external factors	Cause-and-Effect	High	$15,000 - $1+ million	6 months (retrospective) 1-2 years (prospective)	Youth in Action - Colombia National Planning Department (2005)	2
CEA/CBA/SCBA	The impact (non-financial, financial, social) per unit of cost ratio	-	Moderate	+ $10,000 - $30,000	+ 1-3 months	Youth Training Program - Argentina Ministry of Labor (1997)	3

CEA: Cost-Efficiency Analysis; CBA: Cost-Benefit Analysis; SCBA: Social Cost-Benefit Analysis;

[4] Excerpt retrieved from http://www.youth-employment-inventory.org/about/

Evaluator's Necessary Skills for an Evaluation Project

The "M&E Summary" section of this document outlined the main skills that the M&E staff needs to have in order to adequately conduct specific types of evaluation. The table below details these skills requirements.

Skill	Description	Process	Outcome	Impact	CEA/CBA/SCBA
Program Design & Monitoring	* Familiarity with youth livelihood programming * Country knowledge * Experience in program design * University degree in social sciences	Required	Required	Desirable	-
Quantitative Data Collection	* Specialized training in the design and fielding of surveys * Some knowledge of quantitative data analysis * University degree in social sciences	Desirable	Required	Required	-
Quantitative Data Analysis	* Specialized training in statistics or econometrics * A master's or doctorate degree in economics or related field	Desirable	Desirable	Required	Desirable
Qualitative Data Collection	* Specialized training in implementation of qualitative techniques * A master's or doctorate degree in sociology, anthropology, or psychology	Required	Desirable	Desirable	-
Qualitative Data Analysis	* Specialized training in coding and analyzing qualitative data * A master's or doctorate degree in sociology, anthropology, or psychology	Required	Desirable	Desirable	-
Valuation	* Specialized training in estimating the costs and benefits of human service programs * A master's or doctorate degree in economics or related field	-	Desirable	Desirable	Required

This table was adapted from Hempel & Fiala (2011)."

Cost Estimative of Evaluation Projects

The estimated costs for each type of evaluation listed on the Summary Table were retrieved from The World Bank's *"Measuring Success of Youth Livelihood Interventions"* report. They are to be understood as ballpark figures, since the cost of evaluations can vary significantly between different programs and organizations.

Generally, the overall cost of evaluation efforts are mostly influenced by the costs of **acquiring data from programs and their participants** (through surveys, interviews, etc.), which may or may not involve travel. Thus, organizations that already have readily available information may experience lower costs when implementing program evaluations if compared to organizations that are beginning to establish a

performance measurement system[5]. Moreover, the size of the program (in number of participants) and the access to them before, during, and after the intervention are two other factors that can influence the overall cost of the evaluation project significantly.

For most of the evaluation types introduced in this report there are lump sum cost estimates available. However, many organizations around the world estimate their M&E costs by **setting aside 2-5% of the program's budget to M&E purposes**. The International Labour Organization (ILO), for instance, requires that at least 2% of total program funds are reserved for evaluation purposes (International Labour Organization, 2012) and the 5% rule-of-thumb is commonly used by organizations such as UNDP. The table below, retrieved from the World Bank's handbook "Impact Evaluation in Practice" (Gertler, Martinez, Premand, Rawlings, Vermeersch, 2011), illustrates the cost range of impact evaluations—the costlier type of evaluation of the menu of options presented by this report.

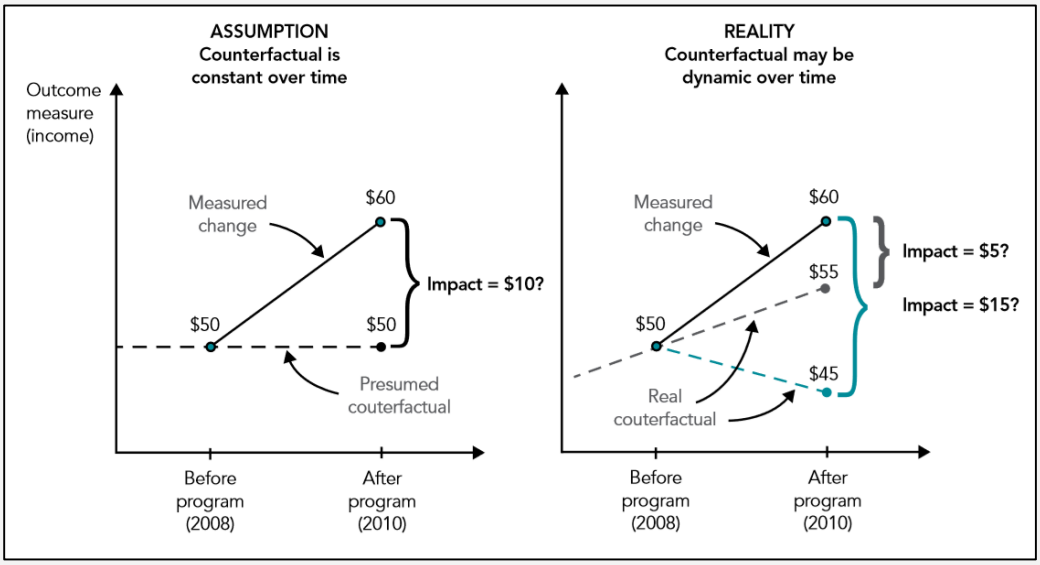
Program Evaluation	Country	Total Cost of Evaluation	Total Cost of Program	Evaluation Cost as % of Program Cost
Migrant Skills Development and Employment	China	220,000 USD	50,000,000 USD	0.4
Social Safety Net Project	Colombia	130,000 USD	86,400,000 USD	0.2
Social Sectors Investment Program	Dominican Republic	600,000 USD	19,400,000 USD	3.1
Social Protection	Jamaica	800,000 USD	40,000,000 USD	2.0
Social Safety Net Technical Assistance	Pakistan	2,000,000 USD	60,000,000 USD	3.3
Social Protection Project	Panama	1,000,000 USD	24,000,000 USD	4.2
1st Community Living Standards	Rwanda	1,000,000 USD	11,000,000 USD	9.1
Social Fund for Development 3	Yemen	2,000,000 USD	15,000,000 USD	13.3
Average	-	968,750 USD	38,225,000 USD	4.5

Since all rows in the table above were impact evaluations of large-scale programs, it is only natural that the cost of the evaluations was also considerably high. However, as a percentage of the program's total cost, the evaluations of the table corroborate with the 5% rule-of-thumb used by many organizations worldwide.

[5] Note that having a performance measurement system in place does not prevent an organization from collecting program-specific data.

Interpreting the Results of Evaluation Projects

Interpreting correctly the results of evaluations, particularly of outcome or impact evaluations, is a crucial requirement for the well-functioning of any M&E system. It is important to recognize that not all evaluations establish **causal relationships** between the program and the change in the outcomes. That is because the some evaluations rely on a before-and-after comparison of the outcomes of program participants (e.g., income prior and after completion of program) without making use of a control group. In these cases, to say that the change in outcomes is attributable to program only is to **assume** that the outcomes would not have changed had the individuals not participated in the program. But what if it was the case, for instance, that one program participant have gotten a well-paying job even without being part of the program? In that case, the before-and-after comparison would be **overstating** the actual effect of the program. The figure on the left-hand side below describes what a before-and-after comparison of outcomes would assume the program's impact to be on a certain participant's income and compares it with the figure on the right-hand side, which shows that **the actual program's impact is contingent on what the income of the program participant would be had he not participated in the program** (also called the *counterfactual*).



Figures retrieved from Hempel & Fiala (2011).

The right-hand side graph shows that measuring a $10 increase in the income of a program participant after the completion of the program may in fact mean $5 impact or a $15 impact, depending on how much the income of the same participant would be had he not participated in the program (the counterfactual).

Naturally, the counterfactual cannot be observed, since it is a hypothetical scenario. It can, however, be estimated through an impact evaluation—a costlier and more time-demanding evaluation project that is recommended only to programs of a certain scale. Part II of this report, which details each evaluation type, explains how the estimation of the counterfactual is done and the basics of impact evaluations.

It is important to emphasize that not only impact evaluations can estimate the counterfactual, although they are, by far, the type of evaluation that most frequently does so. An organization may carry out an

outcome evaluation and make use of control groups to estimate the counterfactual, and thus, be able to establish causality between program and change in outcomes. However, since outcome evaluations are concerned with **short-term outcomes** and impact evaluations are concerned with **long-term outcomes**, impact evaluations are **required** to estimate the counterfactual in order to make sure that external factors are not skewing the results. For example, external factors, such as the country's economic growth rates, are certain to influence more whether or not a program participant has held a quality job for one year (long-term outcome) than whether or not a program participant learned the skills taught by a vocational training program (short-term outcome). Therefore, while both types of evaluation can make use of control groups, estimate the counterfactual, and establish causality, there is more incentive for impact evaluations to do so than outcome evaluations.

Part II

Types of Evaluation

Building M&E Capacity

Part II: Types of Evaluation

Purpose of Part II

This section takes a closer look into the five types of evaluation projects briefly described in Part I. It provides details on the purposes, resource requirements, expected duration, and limitations of each evaluation. Part II also pays special attention to impact evaluations and how they are able to establish causality between a program and its results.

The "additional information" subsections contain practical information about each evaluation type, but it should not be seen as a thorough guide on how to implement an evaluation. In other words, Part II does not replace the need of having an M&E professional in order to carry out a particular program evaluation.

This is a stand-alone part of the report and may be read separately from the rest of it. Thus, because of its level of detail, **it is recommended that the reader refers to Part II when s/he desires to have more information about a particular type of evaluation**.

The framework of analysis that is the basis of Part II is an adaptation of the framework used in the World Bank's report titled "*Measuring Success of Youth Livelihood Interventions*" (Hempel & Fiala, 2011). Adaptations were made to better fit SFSD's context as well as other sources that were used in this report.

Note on Cost Estimations

This report reproduces the cost estimations done by the *Global Partnership for Youth Employment* for each type of evaluation. The estimations are a range of U.S dollar values that does not take into account the size or type of the program—only the type of evaluation being carried out. Therefore, the estimations should be understood as "ballpark figures" and as an indication of the amount of efforts that each type of evaluation entail (meaning that costlier evaluations require more efforts and resources than cheaper evaluations). If SFSD's staff wishes to take into account the size of the program in the estimation, they may set aside 5% of the program's total budget for evaluation purposes. This is a common way of roughly estimating the costs of program evaluations.

More precise estimates can only be produced on a case-by-case basis.

Introduction

Descriptive, Normative & Cause-and-Effect Evaluations

As described in Part I, this report divides the findings of program evaluation into three categories: descriptive, normative, and cause-and-effect (Hempel & Fiala, 2011).

Descriptive evaluations are the most basic form of evaluation and they simply document the activities of a program without making judgements of value—that is, without comparing what has been documented with pre-established goals.

Normative evaluations, besides documenting the activities and/or results of a program, pre-establish goals to evaluate its merits. This is a common type of evaluation since most organizations carry out program evaluations in order to find out the merits and shortcomings of their portfolio.

Cause-and-Effect evaluations go a step further than their normative counterparts. Besides comparing the findings with pre-established goals, cause-and-effect evaluations estimate the hypothetical scenario of the intervention *not* being implemented (counterfactual) so that the program's actual effects can be segregated from the effects of external factors. For instance, the same vocational training program that is implemented during an economic recession as opposed to a growth period may yield significantly different results, although the program itself is the same. Therefore, cause-and-effect evaluations, when done correctly, can comfortably claim that the impacts observed are indeed results from the program.

Monitoring - Descriptive

What does it tell us?

Monitoring projects are the simplest type of evaluation since they are **purely descriptive**. Monitoring consists in collecting specific information about a program through observation and reporting it at a later stage. Thus, a monitoring evaluation tells the organization **what activities took place, who conducted the activities, and who was reached by the activities.** However, the project does not go beyond reporting the information above. As a rough comparison, a monitoring evaluation can be understood as a simplified *process evaluation*.

Monitoring evaluations do not evaluate the outcomes of a program.

Estimated Duration: Unspecified

Since monitoring evaluations are entirely dependent on the observation of the activities of a program, it is difficult to estimate its average duration.

Organizational Capacity: Low

Organizational capacity is the combination of financial, human, and technology resources that an organization must have in order to carry out successfully an evaluation. In the context of this report, monitoring evaluations demand a low organizational capacity.

Financial Resources

As with the estimated duration, the financial resources required for a monitoring evaluations are entirely contingent on the program's characteristics—thus, this report provides no estimate of costs. However, the **M&E staff salary** and the **scope of evaluation** can be considered the main cost drivers.

Human Resources

Part I of this report lists the following skills as necessary for the M&E staff to possess in order to conduct monitoring evaluations:

Program design and monitoring skills, such as:

- Familiarity with youth livelihood programming;
- Experience in program design;
- General knowledge of qualitative and quantitative data collection techniques;

- Country knowledge;
- A university degree in social sciences.

Technology Resources

The technology resources that a monitoring evaluation requires are contingent on the program's characteristics and this report refrains from making specific recommendations. However, since monitoring consists of **observing the activities** of a program, one can comfortably expect that qualitative information collection tools will be necessary. However, it must be noted qualitative information *analysis* is not part of monitoring project, and, therefore, softwares such as NVivo are not required for this type of evaluation.

Process Evaluation - Normative

What does it tell us?

Process evaluations document whether a program has been implemented as intended and the reasons for proper implementation or non-implementation. Process evaluations examine **whether the activities are taking place**, **who is conducting the activities**, **who is reached through the activities**, and **whether sufficient inputs have been allocated**. This may include contrasting actual and planned performance along all or some of the following criteria:

- *The locale where services or programs are provided (e.g., rural, urban);*
- *The number of people receiving services;*
- *The economic status and racial/ethnic background of people receiving services;*
- *The quality of services;*
- *The actual events that occur while the services are delivered;*
- *The amount of money the project is using;*
- *The number of activities and meetings;*
- *The number of training sessions conducted.*

Developing a process evaluation plan can provide the framework and serve as an important first step in an evaluation effort. Process evaluations can be both *formative* and *summative*: **formative** in that results from the evaluation can be used for **program improvement** and **summative** in the sense that evaluation findings can be used to provide **judgment on whether the program is being implemented effectively**.

Process evaluation projects are commonly used to confirm that the program was indeed implemented as intended before using resources to assess its effectiveness.

Estimated Duration: 1-6 Months

The World Bank estimates that process evaluations typically take 1-6 months to be completed (Hempel & Fiala, 2011). Naturally, this period may vary depending on the program as well as on the organization's readiness to carry out an evaluation.

Process evaluations are typically enacted in the following stages of programs:

- *During the early or 'pilot testing' stage of the program to identify and overcome challenges to program delivery;*
- *On an ongoing basis to maintain or improve the quality of the delivery process;*

- *When the program is undergoing major restructuring to "benchmark" current program operations against those that are regarded better;*
- *In conjunction with an outcome evaluation project to understand why a particular program failed or succeeded.*

Organizational Capacity: Moderate

Organizational capacity is the combination of financial, human, and technology resources that an organization must have in order to carry out successfully an evaluation. In the context of this report, process evaluations demand a moderate organizational capacity.

Financial Resources

The main drivers of the cost of a process evaluation are the **salary of the M&E staff** and the **scope of the evaluation**. Due to its qualitative character, process evaluations can vary widely in scope and more complex process evaluation can end up requiring significantly more data collection, warranting travel to distant locations, lengthy interviews and analysis, etc. The estimates process evaluations to **typically cost from 10,000 USD to 60,000 USD** (Hempel & Fiala, 2011). However, process evaluations that are conducted on an ongoing basis (as opposed to a one-time project) may involves more time and financial resources.

Human Resources

Part I of this report lists the following skills as necessary for the M&E staff to possess in order to conduct process evaluations:

Program design and monitoring skills, such as:

- Familiarity with youth livelihood programming;
- Experience in program design;
- General knowledge of qualitative and quantitative data collection techniques;
- Country knowledge;
- A university degree in social sciences.

Qualitative data collection skills, such as:

- Specialized training in implementation of qualitative techniques;
- A master's or doctorate degree in sociology, anthropology, psychology, or other relevant fields.

Qualitative data analysis skills, such as:

- Specialized training in coding and analyzing qualitative data;
- A master's or doctorate degree in sociology, anthropology, psychology, or other relevant fields.

Technology Resources

Process evaluations often utilize both qualitative (questionnaires, interviews, focus groups, etc.) and quantitative (surveys, existing records, etc.) data collection methods. For that reason, SFSD's evaluators may need to secure licenses of programs such as NVivo, MaxDQA, etc. for the type of qualitative analysis that a process evaluation entails. For quantitative analysis, survey softwares such as Qualtrics (for online surveys) and relational databases, such as Microsoft Access or Excel, may be necessary.

Additional Information

Steps in process evaluation[1]

i. Form a working group of key stakeholders to determine whether to bring an outside evaluator or have a staff conduct the evaluation.
ii. Develop the logic model: logic model is a visual representation of how the program is expected to work—connecting resources, activities and the intended changes or impacts the program is expected to achieve. Logic modelling subsumes aspects of describing the program (purpose, underlying theory, strategies and expected outcomes) and describing complete and acceptable delivery of the program.
iii. Determine the audience for the evaluation: external audiences include funders, potential volunteers, community advocates, etc. Internal audience includes staff members, managers and board of directors.
iv. Determine the evaluation questions: a preliminary step in determining the evaluation questions involves determining the purpose of the evaluation project. The evaluation purpose will determine the kind of questions that need to be asked as well as the methods that will be used to answer those questions.
v. Choose the evaluation design and data collection methods. The timing, sources of data, tools and analysis methods are determined following the finalization of the evaluation questions. Data collection for process evaluation can range from direct observations and surveys to in-depth interviews and focus groups.

Example of a Process Evaluation

The Youth Offender Demonstration project initiated by the U.S Department of Labor (DOL) in 1999 started as a 24 month-long pilot project to prepare and place "at risk" youth into employment, thereby breaking the cycle of crime and dependency on public support. The U.S. DOL commissioned a process evaluation to "document the implementation process of the nine Round Two projects over the duration of the demonstration". The evaluation was also expected to document the achievements and challenges in the delivery of integrated services to the youth.

The focus of the process evaluation in this case was mainly formative meaning that the evaluation would provide feedback on improvement of service delivery to grantees, technical assistance specialists, and DOL project officers.

Their process evaluation design followed the Context-Inputs-Process-Product (CIPP) model proposed by Stufflebeam and Shinkfield which involved an assessment of the program in relation to the following questions:

- *Were important needs addressed?*
- *Was an appropriate program design employed?*
- *Was the design well executed?*
- *Was the program successful?*

Based on stakeholder needs and interests, 10 evaluation questions were proposed, each addressing a component of the CIPP model. For instance the evaluation question "What is the context of each project and how did it influence project development and implementation" addressed the context component.

[1] Saunder, Evans, & Joshi, 2005.

The evaluation question "What methods of staff recruitment and training were used and how successful were they", in addition to three other evaluation questions, addressed the input component. The process component was addressed by the 3 evaluation questions including "What types of training, employment and gang suppression programs were provided to the target population and what were the intensity, duration and quality of those programs?" The product component was addressed by the question "what steps have been taken to assure the continuation of integrated services and activities after the project funding ends and what is the likelihood of success?"

The evaluation used a range of data sources: observations, unstructured interviews, document reviews, etc. The analysis tasks for the evaluation involved:

- *Describing the project and how it was implemented*
- *Comparing sites to determine commonalities, differences, barriers and successes*
- *Making a determination as to whether the program was implemented as planned*

Reference Document: Youth Offender Demonstration Project Process Evaluation[2]

Outcome Evaluation - Normative

What does it tell us?

Outcome evaluation seeks to **gauge the extent of success in achieving specific short-term outcomes and the reasons for achievement or non-achievement of such outcomes.** It also validates the contribution of a specific program to the outcomes and identifies key lessons learned and recommendations to improve performance.

Outcome evaluations usually focus on the **immediate effects that the program has in its participants**. Its purpose is to understand intended changes in **knowledge, skills, attitudes, practices, and policies** that result from a program. Assessment of outcomes is an important first step in determining the long-term impact of the program on the society. The findings of an outcome evaluation, along with existing theory and research, can help in determining the potential long-term impact of the program on the broader society. For instance, programs which provide vocational training to women are often grounded on existing theory and research that suggests that financial independence of women has positive ripple effects on the household. If an outcome evaluation of such program attests that financial independence is being fostered, the organization may **suggest** that positive ripple effects on the household are also likely to take place. This is important because long-term impact on society are often the motivation of social programs.

A good outcome evaluation often depends on a thorough process evaluation. An impact evaluation often depends upon well planned process and outcome evaluations.

A program that is backed by sound theory and research also helps with the fact that most outcome evaluations do not use control groups in order to account for the effect of external factors. While existing theory and research alone cannot replace an experimental design[3], it allows for the sponsoring organization to be more secure in regards to the long-term effects of its program.

[2] Miller & MacGillivray, 2002
[3] See Impact Evaluation section for more information

As in the case of process evaluations, **determining the goal or purpose of the evaluation is important when conducting an outcome evaluation**. The outcomes that the program defines through components such as *logframes* or *logic models* become the primary focus of an outcome evaluation. These outcomes must be **observable** and **measurable** and must be of interest to the stakeholders-staff, partner organizations, funders, etc. In cases of the outcomes themselves cannot be observed, SFSD may make use proxies (or *outcome indicators*), which are specific rubrics used to measure each targeted outcome. For instance, self-esteem is an outcome that is not directly observable or measurable, but with the use of rubrics such as the "Rosenberg's self-esteem scale[4]", the evaluators may have a good enough approximation of it in order to carry out an outcome evaluation.

Estimated Duration: 1-3 months

The World Bank estimates that outcome evaluations typically take 1-3 months to be completed (Hempel & Fiala, 2011). Naturally, this period may vary depending on the program as well as on the organization's readiness to carry out an evaluation project.

Organizational Capacity: Moderate

Organizational capacity is the combination of financial, human, and technology resources that an organization must have in order to carry out successfully an evaluation. In the context of this report, outcome evaluations demand a moderate organizational capacity.

Financial Resources

The main cost drivers of an outcome evaluation are the **salary of the M&E staff** and the **scope of the evaluation**. The latter is a main cost driver since it determines the type of information that needs to be collected and, thus, its costs. For instance, if the evaluation seeks to examine how many participants of a vocational training program are employed full-time after the completion of the program, phone calls or electronic surveys may suffice. However, if the outcome evaluation is concerned with whether or not the participants of a disease prevention program changed their behavior at home, more elaborate ways of collecting information—such as in-person interviews that require travel—may be necessary and, therefore, increase the data collection costs.

The World Bank estimates outcome evaluations to **typically cost from 10,000 USD to 30,000 USD** (Hempel & Fiala, 2011).

Human Resources

Part I of this report lists the following skills as necessary for the M&E staff to possess in order to conduct outcome evaluations:

Program design and monitoring skills, such as:

- Familiarity with youth livelihood programming;
- Experience in program design;
- General knowledge of qualitative and quantitative data collection techniques;
- Country knowledge;

[4] The Rosenberg's self-esteem scale asks 10 questions and ranks the responses in order to quantify an individual's level of self-esteem.

- A university degree in social sciences.

Quantitative data collection skills, such as:

- Specialized training in the design and fielding of surveys;
- Some knowledge of quantitative data analysis using Microsoft Excel or similar software;
- Program management skills to build and lead a team of enumerators;
- University degree in social sciences.

Technology Resources

Outcome evaluations that utilize both qualitative (questionnaires, interviews, focus groups) and quantitative (surveys, existing records) data collection methods may require more resources than a purely quantitative outcome evaluation. If qualitative information is part of the outcome evaluation, SFSD's evaluators may need to secure licenses of programs such as NVivo, MaxDQA, etc., which allow for interview and document analysis. For quantitative-only outcome evaluations, survey softwares such as Qualtrics (for online surveys) and relational databases, such as Microsoft Access or Excel, may be necessary.

Additional Information

Steps in outcome evaluation:

i. Develop a theory of action, which is a causal chain of events that ends with the program's intended outcomes (a logic model that describes the short term, intermediate and long term outcomes can also be used as a starting point);
ii. Determine which outcomes are of primary importance to the stakeholders;
iii. Determine *outcome indicators* (proxies) to be measured;
iv. Develop a measurement strategy identifying the methods and instruments to use.

Common outcome evaluation designs:

Before-and-after comparison: conducting pre- and post-intervention assessment is the most commonly used design for capturing program outcomes. It entails collecting information at the onset and closing of the program.

Reflective strategy: participants are asked to reflect back to a baseline level and answer open and close ended questions on what has changed and what caused the change. This design is used when baseline information was not collected at the program start.

Descriptive strategy: outcomes are captured in a narrative manner from the perspectives of the customers and the providers using stories or mini-cases.

Example of an Outcome Evaluation

To address the issue of youth unemployment in **Kenya**, **Microsoft** and the **International Youth Foundation** launched the **Youth Empowerment Program** (YEP) in **2007** to provide "demand-driven training in information and communications technology (ICT), life skills and entrepreneurship" to youth aged 16-35. An independent evaluation of the YEP in Kenya was conducted by FocusAfrica—an independent consulting firm based in Senegal. The evaluation study described here focuses on the evaluation of the program implemented by Nairobits—a nonprofit Kenyan youth organization founded in 1999.

The evaluation was conducted to:

a) Review the implementation process and its outcomes;
b) Assess the program outcomes in terms of acquisition of skills, placement and creation of businesses;
c) Gather opinions of employers on the performance of trainees and employees as well as gaps to be filled; and
d) Gather recommendations from employers and stakeholders to improve the implementation of the program.

The YEP program under Nairobits had the following objectives:

- Train 300 young people between the ages 16-35 in ICT, entrepreneurship and life skills;
- Placement of at least 62% participants, divided among IT jobs, self-employment, and community activities.

Under the outcomes measurement plan, **participants were surveyed on three stages**: at baseline (prior to program start), at program completion, and approximately 6 months after completing the program. The baseline and exit surveys (at program completion) mainly included participants' self-assessment in their life skills abilities, such as: **personal self-confidence**, **conflict management**, **ethics in workplace**, **working in groups**, among others.

In addition to these surveys, the evaluators administered a questionnaire to a sample of program participants to collect information, such as whether the employment generated by the program was sustainable, as well as to capture the overall perception of the program. Additional data collection methods used included: employer questionnaires, participant focus groups, and stakeholder meetings. The assessment focused on placement analysis (overall placement rate, specific activities undertaken by participants, current employment status, etc.), life skills self-evaluation, quality of life, quality of training, participants' perception of the future, and employer feedback on the program.

Reference document: Youth Empowerment Program Evaluation Report (NairoBits, 2010)

Impact Evaluation – Cause-and-Effect

What does it tell us?

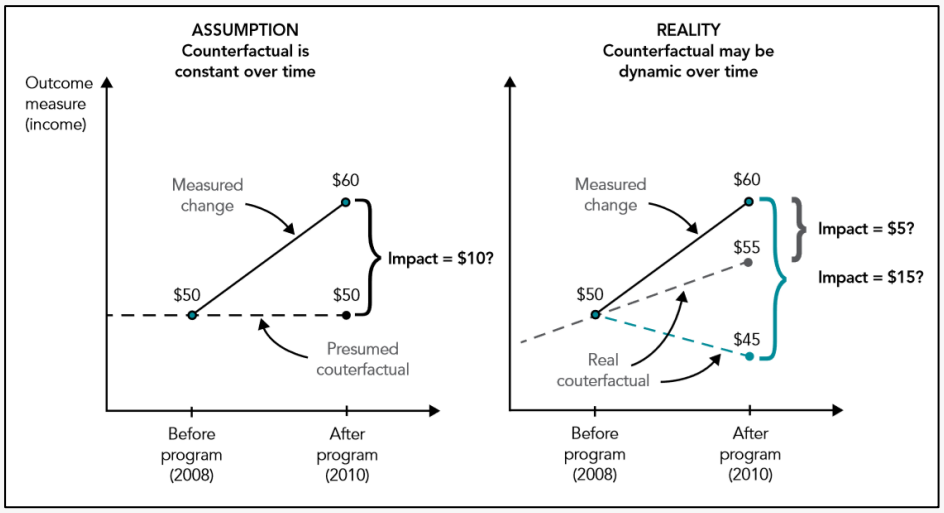
An impact evaluation (IE) attempts to **determine to what extent a program impacts the long-term outcomes of interest**. In other words, a properly done IE establishes causality between the program (e.g., job training) and the end results (e.g., job creation) and measures the extent to which one affects the other. However, IE is also the most demanding type of evaluation, requiring a good amount of financial resources, technical expertise, programmatic flexibility, planning, and time from the implementing organization.

In order to establish causality, impact evaluations estimate the **counterfactual**, that is, **what would have happened if the program was not implemented**, and compares it with the actual outcome (e.g., how many jobs were in fact created). The difference between the counterfactual and the outcome is called the **impact** of the program, which is the **change in the outcome that is attributable to the program itself and not to exogenous factors**.

Why do we need to estimate the counterfactual?

While it may sound intuitive to just compare the outcomes of program participants before a program is implemented with the outcomes after the program is complete, this technique does not yield the true impact of the program. That is because the outcomes of the program participants could change even in the absence of the program. For instance, detecting that the income of the participants of a program increased in comparison to their income prior to the program completion does not consider the **possibility that the participants' income could have increased even if the program was not implemented**. Perhaps some of the participants would have found better jobs even when not receiving the training—one cannot assume that the outcomes would have remained constant over time. **Thus, comparing before and after outcomes, without estimating a counterfactual, is likely to either underestimate or overestimate the impact of the program**.

The figures below exemplify the shortcomings of a simple before-and-after outcome comparison. The graph on the left illustrates a scenario where there is an assumption that the income of a program participant would not have changed if the program had not been implemented. If that was true, the impact of the program would have been of $10 increase, since that was the measured change in income before and after the program was complete. However, the graph on the right shows that, if such assumption is not true and, say, the participant's income would have decreased $5 in case the program had not been implemented, then the actual impact of the program would be of $15. Simply comparing the income before and after the program completion in this case would have underestimated the program's impact by $5.



Figures retrieved from Hempel & Fiala (2011).

Naturally, **the counterfactual cannot be observed**, since it is a hypothetical scenario. It can, however, be estimated. Please refer to the "Additional Information" subsection for more information on how to estimate the counterfactual.

Estimated Duration: 6 – 24 months

The World Bank estimates that **prospective impact evaluations** (of programs that will be implemented) typically take from 12 – 24 months, while **retrospective impact evaluations** (of completed programs) may be done in a period of 6 months. It is worth noting, however, that retrospective evaluations can only be carried out when the program design allows it (for instance, if certain baseline information of program participants was not collected, it may be that the program is not suitable for an impact evaluation).

It is always advisable that impact evaluations are planned for in advance. This type of evaluation is entirely dependent on the program design and embedding M&E in the program's design can greatly enhance the validity of the findings of an impact evaluation project (as well as diminish its costs).

Organizational Capacity: High

Organizational capacity is the combination of financial, human, and technology resources that an organization must have in order to carry out successfully an evaluation. In the context of this report, impact evaluations demand a high organizational capacity.

Financial Resources

The main cost drivers of an impact evaluation are its **extensive, rigorous data collection** and the **M&E staff** (or expert consultants') salary. That is the case because impact evaluations often require several surveys (baseline survey, exit survey, follow-up survey, etc.) of program participants as well as of the control group[5]. Moreover, impact evaluations can only be effectively applied on large programs (the evaluation project needs a large sample size), and thus, data collection costs are a relevant factor. The type of expertise required from the M&E staff also comes at a high financial cost (perhaps for that reason many organizations commission impact evaluation studies as opposed to having full-time impact evaluators on staff).

The World Bank estimates impact evaluations to **typically cost from 15,000 USD to over 1,000,000 USD** (Hempel & Fiala, 2011). The costs of impact evaluations are often a restrictive factor for many organizations and it is not surprising that their findings are highly prized within the international development industry. The table below illustrates the cost range of impact evaluation projects (Gertler et al., 2011)

Evaluation Project	Country	Total Cost of Evaluation	Total Cost of Program	Evaluation Cost as % of Program Cost
Migrant Skills Development and Employment	China	220,000 USD	50,000,000 USD	0.4
Social Safety Net Project	Colombia	130,000 USD	86,400,000 USD	0.2
Social Sectors Investment Program	Dominican Republic	600,000 USD	19,400,000 USD	3.1
Social Protection	Jamaica	800,000 USD	40,000,000 USD	2.0
Social Safety Net Technical Assistance	Pakistan	2,000,000 USD	60,000,000 USD	3.3
Social Protection Project	Panama	1,000,000 USD	24,000,000 USD	4.2
1st Community Living Standards	Rwanda	1,000,000 USD	11,000,000 USD	9.1
Social Fund for Development 3	Yemen	2,000,000 USD	15,000,000 USD	13.3
Average	-	968,750 USD	38,225,000 USD	4.5

[5] See "Additional Information" subsection for more information on control groups.

Human Resources

Part I lists the following skills as necessary for the M&E staff in order to conduct impact evaluations:

Quantitative data collection skills, such as:

- Specialized training in the design and fielding of surveys;
- Some knowledge of quantitative data analysis using Microsoft Excel or similar software;
- Program management skills to build and lead a team of enumerators;
- A university degree in social sciences.

Quantitative data analysis skills, such as:

- Specialized training in statistics or econometrics;
- A master's or doctorate degree in economics or related field.

Technology Resources

Impact evaluations are very demanding in terms of technology resources. **Besides the tools to collect the necessary information**, such as survey softwares like Qualtrics, **impact evaluations require statistical data analysis software** like STATA, SAS, or other equivalent programs. Most programs sell licenses for limited periods of time for a reduced cost, in case impact evaluations are not expected to occur often enough to merit a lifelong license.

Additional Information

There are several methods available for carrying out an impact evaluation (IE), but they all share the need for a **control/comparison group**. A control group is a group of individuals that are as similar as possible to the program participants (often called the **treatment group**) but who do not participate in the program. However, the implementing organization collects information of the control group to use in the evaluation. It is important to note that the control and treatment group should be similar not only in their **observable characteristics**, but also in terms of their **unobservable characteristics**. For instance, two individuals of the same *age*, *gender*, and *annual income* are similar in terms of their observable characteristics, since these are verifiable. However, the same two individuals may be radically different in terms of their *motivation to find a job*—a characteristic that is likely to matter as to whether or not they will be employed in the future but that is not observable. Ignoring the unobservable characteristics of individuals may lead to biased results and compromise the IE. If the control and treatment groups are chosen in a way that creates **systematic, significant differences** between them, the IE suffers from **selection bias**. Selection bias skews the estimates produced by an IE, and thus, invalidates its results. Therefore, the way the control and treatment groups are chosen is key to any IE.

Experimental vs. Non-Experimental Designs

In order to create appropriate control groups, IE experts randomize the selection of individuals that participate in the program and those who do not. The idea is that by assigning a sufficient number of eligible individuals at random to the control and treatment group, the observable and unobservable characteristics of the individuals from each group will be very similar on average. An evaluation that randomizes the selection is considered to have an **experimental design**.

However, randomization is not always possible. For instance, a program may have sufficient budget to serve all eligible members at a specific location, and thus, randomization would require that the program deliberately withholds participation for some of the eligible individuals, which can have ethical implications.

Thus, **quasi-experimental designs** attempt to construct a good enough control group to isolate the impact of the program. Experimental designs, however, are, by definition, more precise than their quasi-experimental counterparts.

Impact Evaluation Methods

There are several methods available to conduct an impact evaluation project. In this section we cover two methods that would perhaps be applicable to SFSD's programs. A complete list of methods can be found in the World Bank's report "*Measuring Success of Youth Livelihood Interventions*".

While each method has different benefits and challenges, the choice is ultimately contingent on the program's characteristics.

Experimental Designs

a. Lottery Design

The lottery design is straightforward: it compares the outcomes of the control group with that of the treatment group. The difference is the impact. However, in order to construct correctly a control and a treatment group, the organization needs to determine the eligibility criteria of the program and randomly select, from the pool of eligible individuals, the same number of individuals for the control and the treatment group.

Example: Suppose an organization will implement a skills training program at a certain location and it has a budget to provide the service to 400 individuals. Individuals that are below the poverty line are considered eligible for the program. The organization then identifies that there are 1,000 individuals that are below the poverty line at the location of implementation. To make use of the lottery design, the organization then needs to randomly select, from the 1,000 eligible individuals, the 400 participants (treatment group) and other 400 individuals for the control group.

The outcomes information is collected from each group and the mean difference in the outcomes of the treatment and control group corresponds to the average impact of the program.

Benefits & Challenges of Using the Lottery Design[6]

↑ Most credible impact evaluation method
↑ Simple design & easy to implement
↑ Analytically simple (does not require advanced statistical techniques)

↓ Requires a control group that, albeit eligible for the program, does not participate in it
↓ Random assignment is only effective if it is maintained throughout the entire program (e.g., selected participants may not drop out, job training information leaks to control group, etc.)

When To Use Lottery Design?

The Lottery Design is ideal to programs that plan for the impact evaluation before their implementation (so that the randomization can be carried out properly). It is also adequate for programs that do not have the

[6] Hempel & Fiala, 2011

budget/capacity to serve the entire eligible population (and thus, mitigating the ethical implications of withholding the provision of services to the control group).

Non-Experimental Designs

b. Difference-in-Difference Design

For some programs, it is not desirable or not possible to randomize the selection of participants. For instance, when funding a program at a certain community, SFSD may have better access to certain neighborhoods over others, and thus, choose to work at the better-suited neighborhoods as a means to reduce implementation costs. In that case, the selection into the program is not random, since there could be systematic differences among the eligible neighborhoods that are not being accounted for by SFSD's selection of where to implement the project.

The difference-in-difference method requires the evaluators to identify a control group that is **as similar as possible** to the treatment group. In the example above, that would mean selecting neighborhoods that are the most similar to the ones where SFSD has good access to and including them in the control group. Then, the evaluators would need to collect information of the control and treatment groups on three different points: a) before the program begins, b) at the beginning of the program, and c) after program is complete. The difference-in-difference method **calculates the change over time in the outcomes of interest for both groups and then compares the difference between the changes that each group experienced** (thus the name, difference-in-difference). The idea is that the difference between the changes that each group experienced is the impact of the program. The graph below illustrates the point:

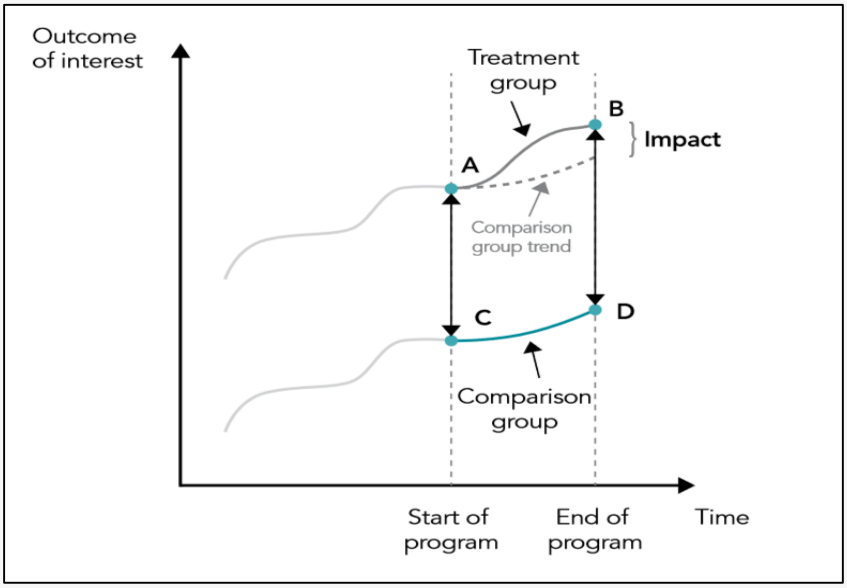


Figure retrieved from The World Bank's "*Measuring Success of Youth Livelihood Interventions (2012)* report.

A = starting level of outcome of interest of treatment group (e.g., income of participants prior to program)
B = final level of outcome of interest of treatment group (e.g., income of participants after program)
C = starting level of outcome of interest of control group (e.g., income of non-participants prior to program)
D = final level of outcome of interest of control group (e.g., income of non-participants after program)

Therefore, the impact of the program is the difference between B and D minus the difference between A and C.

Assumptions of Difference-in-Difference

This method, because it compares the before-and-after of control and treatment groups, accounts for the potential differences in the *unobservable characteristics* between each group. However, it assumes that these differences remain constant over time. That is, while it is not a problem for control and treatment groups to be different in terms of motivation, risk preferences, etc. these differences have to **remain constant over time**. If that assumption does not hold, the difference-in-difference impact evaluation is likely to be either underestimating or overestimating the impact of the program.

Moreover, the difference-in-difference needs a **falsification test**, which is a pre-baseline survey that measures the outcomes of both control and treatment groups, before the baseline survey is carried out. The falsification test is necessary to account for potential differences between treatment and control groups' trends in regards to the outcomes. For instance, if the average income of the treatment group is increasing faster than the average income of the control group, the difference-in-difference method needs to take that into consideration when estimating the impact of the program (otherwise it would be overestimating it).

Benefits & Challenges of Using the Difference-in-Difference Design[7]

↑ Provides a way to account for differences between participants and non-participants
↑ Controls for many individual effects
↑ Can be used retrospectively if the necessary data have already been collected

↓ Produces less reliable results than randomized methods
↓ Requires at least three data collections (two prior to program, one at the end of program), thus potentially increasing the costs of the evaluation project considerably

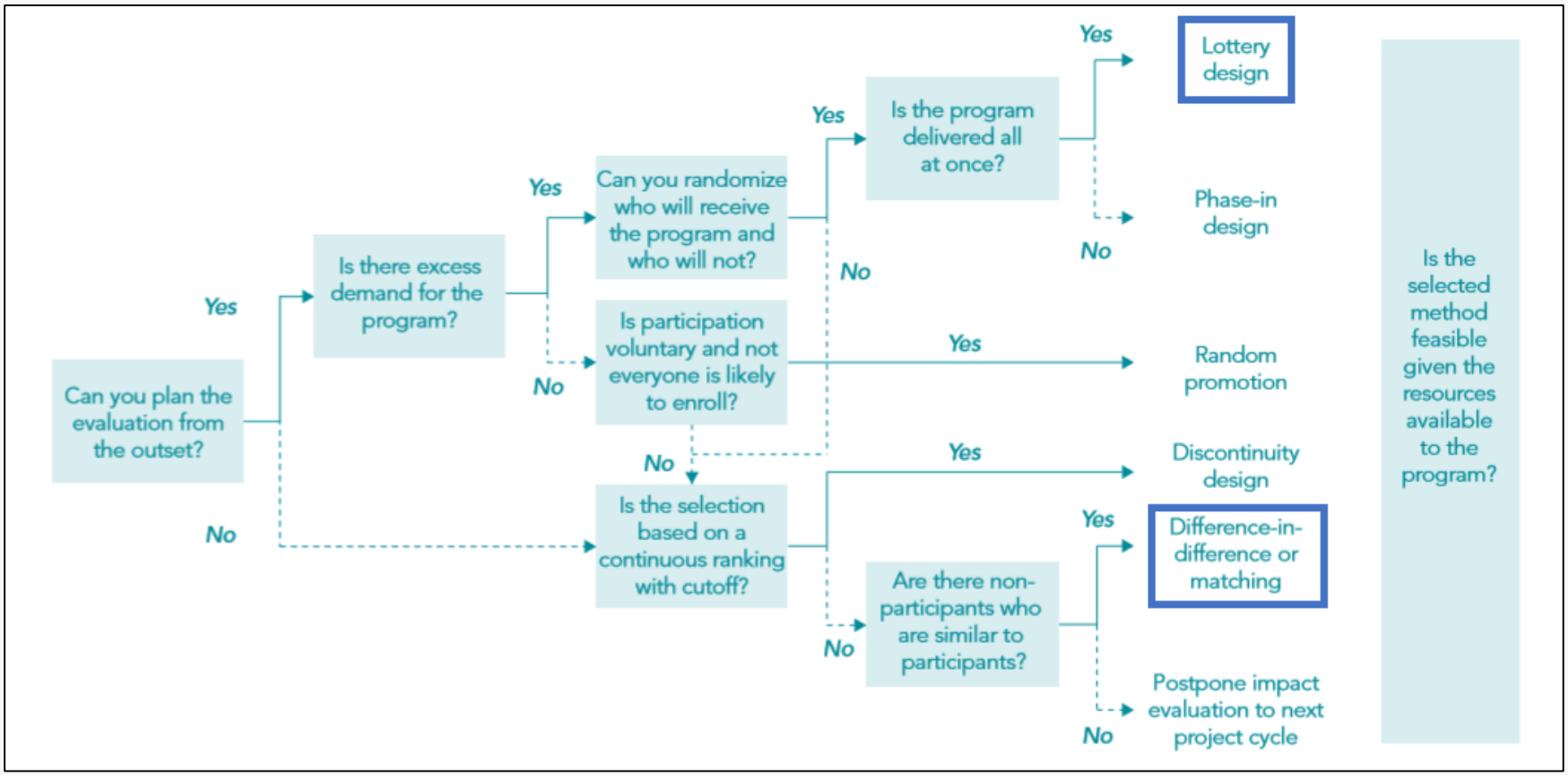
When To Use Difference-in-Difference?

The difference-in-difference is adequate when programs do not have a clear targeting mechanism, which can lead to selection bias. It also can only be used when collecting information of control and treatment group is not a problem, since it requires at least three data collections (two before the program starts, and one upon program completion). For that reason, difference-in-difference is commonly used when there are existing sources of information about eligible individuals, such as household surveys.

[7] Hempel & Fiala, 2011

Choosing the IE Method: Decision Tree

The picture below illustrates the decision process that an evaluator may undergo when choosing the method for an impact evaluation. The two methods covered in this report are highlighted in blue. For further reference on the *phase-in design,* the *random promotion*, the *discontinuity design,* and the *propensity score matching*, please refer to the source document of the table.



Adapted from Hempel & Fiala (2011).

Cost-Effectiveness Analysis, Cost-Benefit Analysis, and Social Cost-Benefit Analysis

What does it tell us?

The Cost-Effectiveness Analysis (CEA), the Cost-Benefit Analysis (CBA), and the Social-Cost Benefit Analysis (SCBA) are complementary types of evaluation that evaluate programs from the financial perspective by looking at the amount (or value) of output generated per unit cost. **The CEA/CBA/SCBA are typically combined with outcome or impact evaluations** so that they not only reveal whether or not the programs are impactful to their target population but that they also make sense from the organization's perspective to be maintained.

Cost-Effectiveness Analysis (CEA)

The CEA aims is to discover whether the program is effective enough to justify its costs. In other words, it seeks to find out whether or not the program is generating enough units of outcomes (e.g. jobs created) per unit cost (e.g. Egyptian pounds).

Cost-Benefit Analysis (CBA)

The CBA is identical to the CEA with the exception that it assigns monetary values to the output of the program. Therefore, instead of measuring jobs created per unit cost, the CBA would measure the monetary benefit for the participant for having a job (e.g., average income increase) per unit cost.

Social-Cost Benefit Analysis (SCBA)

The SCBA is a more encompassing version of the CBA in the sense that it looks at the impact of the program on society and not only on the program participants. For that end, a SCBA project needs to estimate the **shadow prices of the program's inputs and outputs**. Shadow prices are the **prices of abstract commodities** (such as the well-being a community would gain from the construction of a recreational facility) or **the value of a regular good adjusted by societal factors** (such as the regular price of coal plus an additional value that refers to the coal's negative impact on the environment).

Shadow prices also apply to **wages** and **interest rates**, not only commodity prices, which can impact the value of the outputs of a youth employment program. For instance, a vocational training program in a country with high unemployment generates a benefit to society that is not captured by the program participants' salary, but is covered by shadow pricing.

Other than for making use of shadow pricing, a SCBA evaluation is near identical a CBA project.

> Suppose a completed training program that had a total running cost of $200,000 over the course of a year. It trained 200 people and an impact evaluation project revealed that 50% of program participants are employed in the field for which they received training as a result of the program. Thus, it can be said that the program created 100 jobs. Moreover, suppose that the participants who got a job experienced an increase of monthly income of $2,000. However, since only half of participants were employed at the end, it can be said that the program resulted in an average income increase of $1,000/month.
>
> In this example, a CEA would reveal that average costs to create a job is $2,000 ($200,000 divided by 100 jobs created). A CBA would reveal that in order to increase the average monthly income

of program participants by $1, the organization needed to spend $200 on the program ($200,000 divided by $1,000/month average income increase).

A SCBA, on the other hand, would also include the costs and benefits the program brings to the society. These include the externalities and side-effects of the program. These side-effects would need to be monetized and either be added to final ratio of cost per benefit, in case of a negative side effect, or subtracted from the final ratio of cost per benefit, in case of a positive side effect.

Estimated Duration: 1-3 months

The World Bank estimates that CEA/CBA/SCBA evaluations typically take 1-3 months to be completed[8]. The amount and type of additional information about the program that need to be collected for the evaluation project to take place is the most important factor in determining the length of the a CEA/CBA/SCBA.

Organizational Capacity: Moderate

Organizational capacity is the combination of financial, human, and technology resources that an organization must have in order to carry out successfully an evaluation. In the context of this report, CEA/CBA/SCBA demand a moderate organizational capacity.

Financial Resources

The World Bank estimates that a CEA/CBA/SCBA project may cost from **10,000 USD to 30,000 USD** (Hempel & Fiala, 2011) However, **CEA/CBA/SCBAs often function as complementary evaluation** projects that tap into already existing data sources in order to produce their findings, **which has the potential of reducing the cost figures above**.

The main cost drivers of such projects are the **salary of the M&E staff**, which need to have specific valuation and financial analysis skills in order to adequately carry out a CEA/CBA/SCBA.

Human Resources

Part I of this report lists the following skills as necessary for the M&E staff to possess in order to conduct a CEA/CBA/SCBA evaluation:

Valuation skills, such as:

- Specialized training in estimating the costs and benefits of human service program;
- A master's or doctorate degree in economics or related field.

Technology Resources

The technology resources necessary for a CEA/CBA/SCBA are fairly simple. When data collection is not necessary, a spreadsheet program for financial analysis, such as MS Excel, suffices. However, if additional data collection is needed, SFSD may need to secure license of quantitative data collection programs (survey platforms, etc.), depending on the nature of the data that needs to be collected/analyzed.

Additional Information

Examples of information that are typically used in a CEA

[8] Hempel & Fiala, 2011

- The overall costs of inputs of the program;
- The number program participants;
- Outcome measurements (job placement rates, the number of jobs created, etc.);
- A benchmark number of similar program's outcomes.

General steps for conducting a CEA

- Define program goals, inputs (costs), and outputs;
- Quantify the inputs and outputs;
- Calculate the ratio between output and costs;
- Compare program's ratio with selected benchmarks.

Examples of information that are typically used in a CBA

- The overall costs of inputs of the program;
- The number program participants;
- Outcome measurements in their monetary form (monetary benefits to participants);
- A benchmark number of similar program's outcomes.

General steps for conducting a CBA

- Define program goals, inputs (costs), and outputs;
- Quantify the inputs and outputs. Outputs need to be in monetary form;
- Calculate the ratio between output and costs;
- Compare program's ratio with selected benchmarks.

Examples of information that are typically used in a SCBA

- The overall costs of inputs of the program and their shadow prices;
- The number program participants;
- Outcome measurements in their monetary form and their shadow prices;
- A benchmark number of similar program's outcomes.

General steps for conducting a SCBA

- Define program goals, inputs (costs), and outputs;
- Quantify the inputs and outputs. Outputs need to be in monetary form;
- Given the same defined amount of costs, calculate the corresponding benefits;
- Compare the benefits with selected benchmarks.

Assessing Program's Merits

Whether or not the program is worth maintaining, based on the insights that a CEA/CBA/SCBA yields, is ultimately at SFSD's discretion. However, the following three criteria can help the organization's judgement:

- Absolute Criteria: The program's benefits must equal or exceed the impact of giving individuals cash equal to the cost of running the program. Not meeting this criteria indicates inefficiency;
- Relative Criteria: It is better for a program to maintain a benefit level that is no less than the average level of the other similar programs. This shows the program is up to par with comparable programs.

Part III

Case Studies

`

Building M&E Capacity

Part III: Case Studies

Purpose of Part III

This section is a cursory analysis of the M&E system of three selected organizations that implement youth employment training programs in developing countries. The purpose of the analysis is to illustrate, with real world examples, the conceptual framework introduced by Parts I and II of this report.

The section covers each organization's program portfolio as well as their M&E system's advantages, disadvantages, and sample performance measurement system indicators. The data was collected from both academic studies and financial/performance reports from the organizations, besides basic website information.

> Performance Measurement System: In the context of this report, performance measurement system (sometimes referred to just as performance measurement) is a set of tools, practices, and processes that collects, analyzes, and reports key information regarding the performance of an organization. Differently than program evaluation, the performance measurement system compiles information at the organizational level. Indicators that are used in several individual program evaluations (e.g., number of participants employed after program completion) are often made part of the performance measurement system.

About Selection

The three organizations were selected according to following criteria:

a) Program portfolio similar to SFSD's. As mentioned above, all the three organizations provide employment training in developing countries that are, in some ways, analogous to Egypt.

b) Relatively good operational performance. The selected organizations have high job placement rates, considerable operational scale, and sustainable financial budgets.

c) Varied M&E system's characteristics. Instead of selecting well-rounded cases, the organizations were selected based on their different advantages and disadvantages of their M&E systems.

Education for Employment (EFE) - MENA

Program Portfolio

Background

Education for Employment (EFE) is a network of local non-profits launched in 2006. To date, EFE has established branches in Egypt, Jordan, Morocco, Palestine, Tunisia, and Yemen, as well as global offices in the U.S and Europe. The EFE network seeks to help the unemployed youth join the workforce by providing professional and technical training programs.

`

EFE focuses on vocational training, job search skills, and support for early entrepreneurship. It provides job training to over 1,500 young people each year, and the number of trainees are growing. The average job placement rate of EFE has been above 70% in recent years.

Features of Program Portfolio

EFE works with over 400 employer partners and over 50 education partners to **design practical training courses** that meet the employer's' needs. Program participants are trained for immediate employment within a particular industry. EFE's training programs cover a diverse set of skill areas, including **entrepreneurship** training, **hospitality**, **sales**, **textile** industry, and **soft skills for job seeking**. The average program typically takes three months.

M&E Characteristics

Data Collection Methods[1]

EFE has a system of information collection that includes **application forms**, **attendance sheets**, **end-of-training questionnaires, follow-up surveys**, and occasional **focus group interviews**. This allows EFE to conduct a myriad of evaluation projects.

Sample Indicators

- Number of applicants, applicants' level of job-search knowledge and skills, attitudes toward employment, and motivation to get a job (through application forms);
- Number of youth participants enrolled, number of youth who graduate from program, and number of trainings delivered (through attendance sheet and applicant investigation);
- Percentage of graduates satisfied with training, and graduates' self-assessed changes in knowledge and skills (through test and end-of training questionnaire);
- Percentage of graduates applying skills/knowledge, percentage of graduates having found a job, number of recent job applications, and number of recently granted interviews (through short-term follow up survey).

International Youth Foundation (IYF)

Program Portfolio

Background

The International Youth Foundation (IYF) is an NGO founded in 1989. It now has a network in over 98 countries with over 400 partners. IYF cooperates with MasterCard and works to increase employment prospects in both wage and self-employment.

Through its "Passport to Success" program, IYF provides job training to over 10,000 people in diverse programs internationally each year (including MENA). This program, including many subprograms, is implemented in over 40 countries and 20 language areas and is the major program of IYF in employment training[2]. Besides Passport to Success, IYF also have auxiliary programs to enhance the effects of its job training, including alliance building, internships and career building. The average job placement rate is

[1] The data collection methods and sample indicators refer to selected programs of the organization, and not to their entire portfolio.

[2] Passport to Success®: Preparing Young People for the World of Work, 2014

`

over 70%. According to its survey to its former trainees, 75% of its alumni claims that they have benefits from the programs economically.

Features of Program Portfolio

Among other programs, IYF provides programs of **employment training**, **life skills**, **youth leadership**, and **business kick start projects**. Due to its large scale and diversity, the programs are delivered to young beneficiaries through a range of local implementation partners (LIPs). These partners are trained to deliver the individual projects and to grow their capacity for wider delivery. The average training process is about 6 months.

M&E Characteristics

Data Collection Methods[3]

IYF has a system of information collection that includes **exit surveys**, **follow-up surveys**, and **company satisfaction research**, which makes it possible to have both outcome and impact evaluation. Additionally, IYF resorts to **focus groups** whenever qualitative information is necessary.

Sample Indicators

- Number of applicants and applicants' level of job-search knowledge and skills (through application forms);
- Percentage employed in quality jobs or self-employed, percentage who have started loan repayment, percentage in education or training, and percentage reporting self-confidence (through macro digital baseline survey);
- Number of who complete program, number of whom receive job or internship placements, number of who receive small business support, average satisfaction rates, and number of job offers (through exit survey);
- Percentage who retain jobs and business and percentage to re-enroll in education (through post-participation discussions and polling);
- Employer satisfaction with program graduates and participant skills (through employer questionnaire);
- Job retention rate, number of recent job applications and number of recently granted interviews (through post-program focus groups);

Himayat

Program Portfolio

Background

Himayat is a government-sponsored skill development initiative for training and placement of youth in India. It is a part of the Skill Empowerment and Employment scheme of the Prime Minister's Office (PMO). Himayat has its focus on bridging the gap between the industry requirements and the skill set of the youth, so that more employment opportunities can be generated. The project envisages to train and place over 100,000 youths in the next five years.

[3] The data collection methods and sample indicators refer to selected programs of the organization, and not to their entire portfolio.

`

From 2013 to 2014, the Himayat initiative trained approximately 20,000 people per year and its job placement rate is about 80%[4].

Features of Program Portfolio

The Himayat initiative prioritizes training to school and college dropouts. The training is demand-driven, tailoring itself to the needs of the job market. **High-growth service sector occupations** are typically favored. Training is **provided by certified private-sector and nonprofit firms**, which helps to ensure that those providing the training have credible experience. Himayat has put in place incentives to ensure its partners are providing effective training: training firms are paid per student and only if the student is placed in a job upon completion of training and retains that job for a minimum of three months. Himayat training includes **technical skills training** required for particular jobs but also incorporates **soft skills training**. The job training process of Himayat is relatively long and some training process can be up to 1 year.

M&E Characteristics

Data Collection Methods[5]

The Himayat initiative emphasizes information collection through **progress reports**, **employer satisfaction surveys**, **trainee satisfaction surveys** and **training grading systems**. The latter is an inner comparison between its specific training program groups, which makes it possible for it to identify weaknesses.

Sample Indicators

- Number of applicants, applicants' background of job-search knowledge and skills, and attitudes toward employment (through application forms);
- Average satisfaction rates (through satisfaction survey);
- Graduates' self-assessed changes in knowledge and skills, graduation rate, and job placement rate (through exit survey and short-term focus group);
- Employer satisfaction (through employer survey);

[4] Information retrieved from http://www.himayat.org/

[5] The data collection methods and sample indicators refer to selected programs of the organization, and not to their entire portfolio.

`

Information Collection Methods – Summary Table

Information collection methods utilized by the 3 organizations[6]

Information Collection Method	EFE	IYF	Himayat
Application Form	X	X	X
Baseline Survey	-	X	-
Interview or Progress Report	X	-	-
Attendance Sheet or Stand-by Policy	X	-	-
Post-training Test	X	-	-
Exit Survey	X	X	X
Satisfaction Survey	-	-	X
Short-term Follow-up Survey	X	X	X
Long-term Follow-up Survey	X	X	-
Focus Group	X	X	X
Employer Satisfaction Survey	-	X	X
Training Grading	-	-	X
Digital Analysis	-	X	X
Total	8	8	8

[6] The list of data collection methods refer to selected programs of the organization, and not to their entire portfolio.

Part IV

Assessment of Current M&E Capacity

Building M&E Capacity

Part IV: Assessment of Current M&E Capacity

Purpose of Part IV

The purpose of this section is to situate SFSD in the analysis framework introduced by Parts I and II as well as to compare SFSD with the case studies of Part III.

Program Evaluation & Performance Measurement System

Program Evaluation

Evaluation Type	SFSD	QOE
Monitoring	**X**	**0**
Process Evaluation	**-**	**0**
Outcome Evaluation	**-**	**1**
CEA/CBA/SCBA	**-**	**1**
Impact Evaluation	**-**	**2**
CEA/CBA/SCBA	**-**	**3**

In terms of program evaluation, SFSD currently conducts **monitoring evaluations**. While SFSD may not refer to it as such, the act of **verifying the activities** carried out by partner NGOs and of **collecting basic descriptive data** of the programs (such as number of participants enrolled in program) characterizes an unstructured form of monitoring.

It is worth noting that, for not measuring specific outcomes of selected programs, monitoring has a QOE grade of 0, as per the ranking used in the World Bank's Youth Employment Inventory.

Performance Measurement System

Based on conversations with SFSD's staff and on the documents provided, the consulting team identified that the following indicators are currently tracked by SFSD:

- Number of program participants;
- Number of female program participants;

- Participant's self-assessment of soft-skills (such as job search skills);
- Grading of participants at training programs.

The tables below provide a comparison between the indicators and data collection tools used by SFSD and a list of examples that was retrieved from sample programs of the three case studies. The intention of the table is to provide SFSD with a list of potential indicators to include in its performance measurement system as well as to provide insights on what additional indicators could be used for the same purpose.

Possible Indicators for Performance Measurement System

Before program implementation

Indicator	Suggested Data Collection Method	SFSD
Number of program participants	Application forms	X
Number of female participants	Application forms	X
Participant's attitude towards employment	Application forms, baseline survey	X
Motivation to get a job	Application forms, baseline survey	
Participant's self-assessment of job skills	Application forms, baseline survey	
Participant's self-assessment of soft skills	Application forms, baseline survey	X
Participant's self-confidence / self-esteem	Application forms, baseline survey	
Participant's annual/monthly income	Application forms, baseline survey	
Assessment of job skills	Pre-training test	
Assessment of soft skills	Pre-training test	
Number of participants who are employed	Application forms, baseline survey	X

After program implementation

Indicator	Suggested Data Collection Method	SFSD
Number of participants who completed the training	Exit survey	X
Number of female participants who completed the training	Exit survey	X
Participant's attitude towards employment	Exit survey, focus group interviews, short/long-term follow-up survey	X
Motivation to get a job	Exit survey, focus group interviews, short/long-term follow-up survey	
Participant's self-assessment of job skills	Exit survey, focus group interviews, short/long-term follow-up survey	
Participant's self-assessment of soft skills	Exit survey, focus group interviews, short/long-term follow-up survey	X
Participant's self-confidence / self-esteem	Exit survey, focus group interviews, short/long-term follow-up survey	
Participant's annual/monthly income	Short/long-term follow-up survey, focus group interviews	X
Number of training sessions delivered	Visits/observation, progress reports	X
Grading of participants (soft or job skills)	Progress reports, post-training test	X
Number of participants who are applying skills in job	Short/long-term follow-up survey, focus group interviews	
Number of recent interviews granted per participant	Short/long-term follow-up survey, focus group interviews	
Number of job applications per participant	Short/long-term follow-up survey, focus group interviews	
Number of participants who are employed	Short/long-term follow-up survey, focus group interviews	X
Assessment of training quality	Visits/observation, progress reports, exit survey	X
Employer satisfaction with program graduates	Employer satisfaction survey	

** the lists of the indicators and data collection methods by no means encompass all options that SFSD may choose to use in its performance measurement system, but rather serve as a starting point.*

Comparison of Information Collection Methods

Information Collection Method	EFE	IYF	Himayat	SFSD
Application Form	X	X	X	X
Baseline Survey	-	X	-	-
Interview or Progress Report	X	-	-	-
Attendance Sheet or Stand-by Policy	X	-	-	-
Post-training Test	X	-	-	X
Exit Survey	X	X	X	X
Satisfaction Survey	-	-	X	-
Short-term Follow-up Survey	X	X	X	X
Long-term Follow-up Survey	X	X	-	-
Focus Group Interviews	X	X	X	-
Employer Satisfaction Survey	-	X	X	-
Training Grading	-	-	X	X
Visits/Observation	-	-	-	X
Digital Analysis	-	X	X	-

Part V

Designing an Evaluation Plan for SFSD's Projects

Building M&E Capacity

Part V: Designing an Evaluation Plan for SFSD's Projects

Purpose of Part V

The purpose of this section is to present SFSD with an overview of the **steps involved in planning an evaluation**[1]. Additionally, this section includes **an evaluation plan** for one of its projects. Each step has been briefly described with the purpose of providing a basic grasp of the terminologies and questions SFSD may ask (or direct an external evaluator to ask) while developing an evaluation plan for its projects. It is important to note that the descriptions and tools provided under Part V are fairly cursory by nature and each section and sub-section can be a full report in itself. However, the purpose of this part is also to **document the process** utilized in developing the evaluation plan for the *Vocational Productivity Program.*

Introduction

The process of conducting an evaluation involves 4 broad actions:

1. *Focusing the evaluation*
2. *Selecting the evaluation design and the data collection methods*
3. *Collecting and analyzing data*
4. *Communicating and reporting findings*

Part V covers **steps 1 and 2**, often collectively referred to as an **Evaluation Plan**.

Steps in Evaluation Planning

Preparing an evaluation plan is the first step in conducting an evaluation study. In addition to describing what is being evaluated, the plan outlines and reinforces why, how and when it is being evaluated. It functions as an agreed upon document between the evaluator or the person leading the evaluation and the organization that commissioned the evaluation. The process of preparing an evaluation plan is mostly participatory and involves people who are interested in the evaluation findings. In addition to providing direction to the evaluation, the participatory process "increases the understanding of evaluation, the program being evaluated and the use of evaluation results".

> As described by Russ-eft and Preskill (2009), the overall purpose of focusing an evaluation and creating an evaluation plan is to understand the program's underlying assumptions, activities, resources, short- and long-term outcomes and the needs and expectations of the various stakeholders.

Preparing an evaluation plan involves a series of steps:

1. *Drafting an object (evaluand) description.*
2. *Developing the purpose of the evaluation*

[1]The content for the section on steps in evaluation planning has been summarized from the book: *Evaluation in Organizations: A Systematic Approach to Enhancing Learning, Performance, and Change* by Russ-Eft and Preskill (2009)

3. *Developing a program logic model*
4. *Identifying the evaluation's stakeholders*
5. *Determining the evaluation's key questions*
6. *Determining the indicators, data collection sources, methods and timeline*

Object Description

An object description is essentially a detailed description of the program you are evaluating. The information for drafting an object description usually comes from program documents and interviews with staff. It includes a description of:

- Program rationale: why it exists, what need is it serving;
- Program goals or expected outcomes including how it fits into the larger social or environmental change;
- Staffing: key leaders or decision makers, program advocates and program adversaries;
- Budget: funding sources;
- Participants: diversity in terms of race or gender, selection of participants;
- Activities: what services are provided and how;
- Program history: length of time it been operating; significant changes in scope or goals, major changes in the program or its parent/sponsoring organization or funders; and
- Program setting: where are program services delivered.

The description uses a **non-judgmental tone** as it only involves setting up what you are evaluating.

Evaluation Rationale and Purpose

An evaluation can serve the following purposes:

1. Enable decision making: examine whether the program is meeting its goals, whether continued funding should be provided, are mid-course program corrections necessary
2. Enable organizational learning: provide feedback for improvement

The evaluator must be clear about which of these aspects the stakeholders are interested in. Often times, evaluations are used for both decision-making as well as organizational learning. The search for evaluation rationale and purpose begins by asking stakeholders questions such as:

- What do they hope to learn from the evaluation?
- What decisions do they hope to make based on evaluation results?
- When will decisions be made? By whom? When will they be made?
- How much influence the evaluation will have on decision making, realistically?

The common practice is to write down the evaluation purpose statement in two or three concise statements.

The evaluation purpose statement serves as an important reference in making decisions about which evaluation questions must be given priority as opposed to questions that would yield only marginal benefits if answered.

Logic Model

A logic model is a visual depiction of how the program is expected to work, the theory that drives the program and the relationship between the resources, program activities and the short- and long-term outcomes. The use of a logic model started as a practice is the 1960s but came to be used frequently in the 1990s.[2] The **W.K. Kellogg Foundation Logic Model Development Guide** and an online course from the **University of Wisconsin Extension** have for long served as popular resources for the development of a logic model. The creation of a logic model often involves the engagement of stakeholders to understand these relationships and answering the following questions:

1. *What are the assumptions underlying this program?*
2. *What resources (human, financial, organizational) will be used to accomplish this program?*
3. *What activities will be undertaken?*
4. *What direct products (outputs) will provide evidence that the program was implemented?*
5. *What immediate outcomes are expected? What long term outcomes are expected?*

One of the drawbacks of a logic model is that it assumes a linear relationship between the activities and the outcomes of the program.

Most of the social welfare programs operate in a complex socio-political environment and **the intended outcomes do not necessarily follow a linear path** *(see Logic Model for Vocational Productivity)*. While the logic model gives some sense of this complex program theory that drives most development initiatives, it fails to accurately depict the web of linkages between activities and outcomes and between short-, intermediate-, and long-term outcomes. Developing a theory of change (specifically an outcomes chain) is suggested as an alternative to overcome some of these drawbacks. However, a logic model serves as an effective **framework for creating performance measurement points**, **revealing assumptions that underlie the program theory**, **promoting communication regarding the outcomes of the program** and **building understanding among those responsible for the program**.

Stakeholder Analysis

The main premise under which an evaluation is conducted is that it would serve the information needs of different groups of people often referred to as stakeholders.

The extent to which evaluation findings will be utilized hinges on the buy in from the stakeholders involved in the evaluation process

Further, different stakeholders act differently on the findings of an evaluation. Some of them may use it for decision-making while others may use it for understanding the program better. Thus, involving and engaging stakeholders becomes a crucial part of any evaluation study. Stakeholders that wield a decision-making power regarding the program (ex. funders, implementers, etc.) are identified as the primary stakeholders or the **primary users** of the evaluation. **Secondary or tertiary stakeholders** are entities that are

[2] Jodi Sandfort's Lecture on Logic Models

removed from the daily operations and may not have financial or decision-making power related to the program (ex. participants, trainers, parents, etc.). The following criteria are to identify evaluation stakeholders:

1. *Who has a vested interest in the evaluand (program) and the evaluation outcomes?*
2. *Whose position could be affected by the evaluation findings and by the actions taken on the findings?*
3. *Who cares about the program?*
4. *Who has the right to know what the evaluation findings are?*

The identification of evaluation stakeholders and the primary users in particular is important to ensure that the evaluation addresses their concerns or priorities.

Evaluation Questions

Evaluation Questions are broad, overarching questions that guide an evaluation. They form the **boundary of the evaluation** and communicate to the stakeholder **what the evaluation will and won't attend to**. Some of the characteristics of these questions include:

1. *The questions are general, overarching and not so specific that they might be considered interview or survey questions.*
2. *They are **not yes/no** type questions.*
3. *They are prioritized on the basis of what questions are of most immediate concern and what would be "nice to answer". The "nice to answer" questions are either subsumed by the main questions or considered for future evaluations.*

There is no set standard on how many questions an evaluation should answer. However, most evaluations have anywhere between **3-5 questions**. The more questions an evaluation seeks to address the more costly the study will be. It is important to remember that jumping to develop survey or interview questions without developing evaluation questions poses risk that the data collection instruments may not get to the right questions.

Evaluation Design

As defined by Michael Quinn Patton, "program evaluation is the **systematic** collection of information about the activities, characteristics, and results of programs to make judgments about the program, improve or further develop program effectiveness, inform decisions about future programming and/or increase understanding". The first part of this definition highlights the use of a systematic process in conducting evaluations. This leads us to focusing on aspects such as evaluation design and data collection methods. A sound evaluation is informed by a rigorous evaluation design and data collection methods that suit the program context.

The issue of credibility of evaluation findings often raises considerations of data validity. **Internal validity** refers to the extent to which the evaluation correctly answers the questions it claims to answer regarding what is being evaluated. One of the threats to internal validity includes not measuring the effect of confounding variables such as previous knowledge or history on a particular outcome. **External validity** refers to the extent to which the results can be generalized. And **multicultural validity** refers to the relationship between a particular cultural context and the findings of an evaluation. The issue of data

validity gets complicated when the data is qualitative because the "qualitative paradigm assumes that there are multiple realities in the minds of individuals". Establishing credibility for qualitative data often involves using '**triangulation'** that is collecting data from multiple sources or using multiple methods or data collectors or using different perspectives to interpret the data.

We will now turn to identifying some of the common evaluation designs:

- One-shot design: The measurement takes place at one point and is usually done with a post-event survey administered to participants.
- Retrospective pretest design: It is a variation of the one shot design in which the participants report retrospectively on their attitudes, knowledge or skills as well as report on their current attitudes, knowledge or skills.
- One group pretest posttest design: It involves data collection before the intervention as well as following it.
- Posttest only control group design: Two groups are randomly selected, with one group receiving the intervention and the other not receiving the intervention. The two groups are then given the posttest following the intervention.
- Time series design: It involves repeated data collection before and following the intervention.
- Pretest posttest control group design: Two groups are randomly selected with one group receiving the intervention and the other not receiving the intervention. Each group completes a pretest at the same time and after the first group (the group that receives the intervention) completes the intervention, the two groups are administered a posttest.
- Regression discontinuity design: This design involves two groups-one that receives the intervention and one that does not. However, rather than randomly assigning individuals to the two groups, individuals are assigned based on the results of a pretest measure called the *cutoff score* which is used to select individuals most in need of the training or the program. The remaining individuals are assigned to the control group.
- Case study design: This design is particularly useful when asking how and why questions related to individuals, processes or organizations. They use qualitative methods such as individual and focus group interviews, observations and archival records, though they may also employ quantitative methods such as surveys or tests.
- Mixed methods design: This design uses a combination of qualitative and quantitative methods.

Data Collection Methods

The next step in evaluation planning is determining the data collection methods that will be used to gather evaluation data. Some of the factors that determine the choice of data collection methods include:

- Level of acceptable intrusiveness: Some data collection methods like interviews are more intrusive and can conflict with cultural boundaries or may involve interruption of normal business (ex. interviews conducted with professionals during their work hours)

- Resources: Certain data collection methods such as focus groups are more resource intensive as compared to others. For instance, focus groups require multiple facilitators/data collectors whereas surveys can be administered from a single point.
- Evaluation's key questions: A particular data collection method may be required to gather specific kinds of data, for example knowledge retention due to a program may be captured using certain kind of written or oral testing.
- Evaluator skills: Designing and implementing surveys, for example, requires specific training on part of the evaluator.
- Stakeholders' preferred type of data: The choice of data collection method is often driven by the type of data preferred by stakeholders. For instance, stakeholders may want to know program outcomes in the form of success stories. In this case, interviews can be used to capture such data from the participants.
- Instrument validity: It refers to the extent to which the instrument can accurately measure what it intends to measure.
- Instrument reliability: It refers to the consistency with which the instrument can measure what it is supposed to measure if it is used over time.
- Timeliness: It is important to consider the time required to implement certain data collection methods.

The following table provides a list of the different types of data collection methods and their examples.

Data collection methods	Examples
Archival data	Monthly or quarterly reports, training records, employment records
Observation	Qualitative field notes, checklist of activities/behaviors completed by evaluator or program staff
Surveys	Post-program reaction forms, survey of participants and/or employers
Paper or computer based tests	Paper based test at the end of a training or workshop, online test administered a month after completion of the workshop
Individual and focus group interviews	In-person individual or focus groups with participants, telephone interviews with participants

Vocational Productivity Program Evaluation Plan

Overview

The Vocational Productivity (VP) program is being implemented by the Mazallah Foundation for Social Development -a subsidiary of the Aga Khan Cultural Foundation, Egypt. The Aga Khan Cultural Foundation, Egypt has worked on the revival and development of the Darb Al Ahmar community since 2000. The Foundation's work in the community involved working on the revival and development process and the gardening activities in the area. Later on, the Mazallah Foundation for Social Development was created as an independent team that would focus solely on development projects.

The main objective of the VP program is to "raise the standard of living" of the participants and their families and make them self-sufficient through providing them training and jobs in the vocational fields. In line with these objectives, the program aims to focus on bridging the gaps between the skills required by the private sector employers and those that exist among the members of the community currently. The program also intends to connect trainees to employment opportunities in the relevant sectors so as to provide them with an income that meets their needs. The program's focus on training women and providing them with self-employment opportunities aims to overcome the cultural boundaries that prevent them from being financially independent. SFSD also sees the program as acting as a catalyst to build and nurture partnerships between civil society organizations and the private sector in the creation of trained skilled labor.

The program targets 276 beneficiaries in the age group of 18-35 from the Cairo Governorate covering areas of Al Darb Al Amar, Al Gamaleya, Al Sayeda Zaynab, Manshiet Nasser and Al Khalifa. The duration of the project is 18 months and the first round was implemented from May 2014 to October 2015. Currently, the project is undergoing a second round of implementation as the first phase was deemed to be successful.

The main activities of the project involve training the participants in the following skill areas:

- *160 hours of general carpentry training (20 days x 8 hours) per course*
- *56 hours of paint training (7 days x 8 hours) per course*
- *72 hours of leather training (24 days x 3 hours) per course*
- *56 hours of upholstery training (7 days x 8 hours) per course*
- *96 hours of patchwork training (32 days x 3 days) per course*

Throughout its course, the project conducts 5 courses in general carpentry training, 4 courses in paint, leather and upholstery training each and six courses in patchwork training. Each course has 12 participants and the tally of participants trained in each of the trades is as follows:

- *60 in general carpentry*
- *48 each in paint, leather and upholstery*
- *72 in patchwork*

The project undergoes five phases:

- *The first phase lasts for 2 months and involves tasks such as an appointment of the project team, announcement of the project and its objectives, meeting of the facilitation committee, recruitment of trainers, admission of participants, and so on. (The facilitation or steering*

committee is comprised of different stakeholders such as government and private entities coordinates and conducts follow up on the implementation of the project activities).

- *The second phase lasts for about 4 months and involves 2 cycles of paint training, 1 cycle of leather training and 1 cycle of patchwork training followed by graduation of these trainees and a marketing exhibition directed at selling the products of the training.*
- *The third phase lasts for about 3 months and involves 1 cycle of paint, 1 cycle of leather, 2 cycles of patchwork and 1 cycle of upholstery training followed by graduation and a marketing exhibition.*
- *The fourth phase lasts for about 4 months and involves 1 cycle of paint training, 1 cycle of leather training, 2 cycles of patchwork training and 2 cycles of upholstery training followed by graduation of trainees and a marketing exhibition. Other tasks such as meeting of the facilitation committee, delivery of certificates, etc. also take place during this phase.*
- *The fifth and final phase of the project lasts for 4 months and involves 2 cycles of general carpentry training and 1 cycle each of paint, leather, patchwork and upholstery training. Additionally, the implementing agency submits financial and technical reports to the donor and holds a closing conference of the project. Another meeting of the project's facilitation committee is held.*

Phases	Duration	Activities
1	2 months	Project preparation phase
2	4 months	2 cycles of paint training, 1 cycle of leather training, 1 cycle of patchwork training
3	3 months	1 cycle of paint, 1 cycle of leather, 2 cycles of patchwork, 1 cycle of upholstery training
4	4 months	1 cycle of paint training, 1 cycle of leather training, 2 cycles of patchwork training, 2 cycles of upholstery training
5	4 months	2 cycles of general carpentry training, 1 cycle each of paint, leather, patchwork and upholstery training Project closing

Summary of the program phases and activities

The Mazallah Foundation partners with the the Cairo Governorate and training centers such as the Wood Partnership Unit and the Central Cairo Association in the implementation of the project. Additionally, it partners with the agencies such as the EBEX Furniture Company and the Arab Carpentry Factory to guarantee employment to its trainees. SFSD is the main funder contributing nearly 73% of the program's budget. The remaining 23% is contributed by the Mazallah Foundation itself. The total budget for the program is 1,122,300 Egyptian pounds (EP).

Evaluation Purpose

The purpose of this evaluation is to inform SFSD about the quality of the trainings being conducted and how they could be improved, whether the training helps the participants get better jobs, whether the jobs are sustainable, and the effect of the training and subsequent employment on the families of the participants. SFSD is also interested in exploring the influence of the program on aspects such as the psychological well-being of women and the crime level in the community in the long term.

Program: Vocational Productivity program (SFSD) Logic Model

Inputs	Activities	Participation	Outputs	Outcomes		
				Short term	Intermediate	Long term
Funding (SFSD & Mazallah) Employment protocols (ex. EBEX Furniture Company) Training Centers (ex. Wood Partnership Unit, Central Cairo Association) Training staff Training equipment and material Government partners	Project preparation phase: Appoint project team Announce the project Conduct facilitation committee meeting Contract the trainers Obtain employment protocols Training phase: Conduct 160 hours of training in carpentry, 56 of paint training, 72 hours of leather training, 56 hours of upholstery training and 96 hours of patchwork training Post completion phase: Hold marketing exhibition of the products of the training, deliver graduates' certificates, prepare technical and financial reports for donors	276 participants aged 18-35 from the Cairo Governorate (60 in general carpentry, 48 each in paint, leather and upholstery, 72 in patchwork)	**For trainers:** 160 hours of carpentry training provided 56 hours of paint training provided 72 hours of leather training provided 56 hours of upholstery training provided 96 hours of patchwork training provided **For participants:** 100% of participants complete all hours of training 100% of participants acquire certification in carpentry, leather, painting, patchwork and upholstery	Improved knowledge of participants in carpentry, painting, leather, patch work and upholstery Improved skills of participants in carpentry, painting, leather, patch work and upholstery	Increased number of participants secure employment Increased number of participants earn livable wages Increased number of women-owned business	Increased income of participants More number of participants can meet their basic needs Changed community perceptions regarding the participants Increased size of skilled labor force Improved financial status of women Improved confidence and psychological well-being of women participants

Assumptions
Participants are employed in the area they are trained, participants are interested in the training areas, employers find the skills useful, participants are trained in soft skills such as communication, women receive training in business development

External Factors
Country's political climate allows training to be conducted in an uninterrupted manner, political and economic climate of the country allows for job opportunities to be available

Stakeholder Analysis

Stakeholders → / Concerns ↓	SFSD	Mazallah Foundation	Cairo Governorate	Private employers	Trainers	Participants	Community
Do the trainings meet the expectations of the participants?	x	x			x	x	
Are participants satisfied with the quality of the training? • Content • Frequency	x	x			x	x	
Do the skills acquired by the participants meet the needs of the employers?	x	x		x	x	x	
Are the trainers well-qualified in the trade as well as in teaching?	x	x		x		x	
Do the trainings help participants get better jobs?	x	x	x		x	x	
Does the training result in participants securing long term-employment?	x	x	x		x	x	

Does the training and subsequent self-employment opportunity improve the well-being of women? • Financial • Psychological	x	x	x		x	x	x
Does the training and the subsequent employment result in increased income for the participants?	x	x	x		x	x	x
Does the program help reduce crime in the community?	x	x	x		x	x	x
Does the self-employment for women benefit their family especially their children?	x	x	x		x	x	x
Do the trainings improve the knowledge and skills of participants in carpentry, paint, leather, upholstery and patchwork?	x	x		x	x	x	

Key Evaluation Questions

1. *In what ways and to what extent does the program influence:*
   - *Participants and women in particular*
   - *Communities*
2. *To what extent is the program delivered as intended?*

Evaluation Design

The evaluation will serve both formative and summative purposes: formative in that the evaluation findings will be used for program improvement and summative in that the findings will be used to make a judgment about the program. The evaluation will focus on both the process as well as outcomes. The evaluation will use a one-shot design that is the outcomes of the program will be measured at the completion of the program. The advantage of the one-shot design is that it is simple and cost-effective. The disadvantage is that it does not control for other factors such as prior knowledge (for example in carpentry, upholstery, etc.) and the results cannot be generalized. The evaluation will use a mixed-methods approach: it will draw on both qualitative and quantitative data to support the findings.

Evaluation Plan

1. In what ways and to what extent does the program influence:

- Participants and women in particular
- Community

Indicators	Sources	Methods	Timeline
Participants • Change in the knowledge and skills of participants in carpentry, painting, leather, patch work, and upholstery	Program participants (60 participants in general carpentry, 48 each in paint, leather and upholstery, and 72 in patchwork	Skills/knowledge test capturing change in participants' knowledge and skills Expert judgment or grading of products produced as part of the training program	Administered after each cycle followed by aggregation of data after program completion
• Change in the number of participants who are employed (including self-employment) - Duration and type of employment, wages • Change in the income of participants • Ability of participants to meet basic needs	-	Phone interview with participants	Administered 3, 6 *and 12 months* after completion of each cycle
Communities • Change in the perception of community members regarding the participants	Members of the different communities within the Cairo governorate	Focus groups with community members	Conducted 6 months after program completion

2. To what extent is the program delivered as intended?

Indicators	Sources	Methods	Timeline of data collection
Trainers • Total number of trainers for each trade • Qualifications/experience/skills in trade and teaching • Gender distribution of trainers	Program records	Database review	Conducted after program completion
Employer perceptions • Satisfaction with participants' skill quality • Alignment of training content and duration with job requirements • Gaps in implementation • Strengths of the program/implementation	Private employers (ex. EBEX Furniture Company)	Telephone interviews	Administered 3 months after completion of each cycle followed by aggregation of data after program completion
Participants • Participant satisfaction with the training (content, duration, trainer experience and qualifications, location and environment of the training center, etc.) • Barriers to participation • Hours of training received	Program participants	Paper based/online surveys capturing participant satisfaction with the training	Administered after each cycle followed by aggregation of data after program completion
• Gender distribution of participants • Geographic distribution of participants • Age distribution of participants • Attendance rates	Program records	Database review	Aggregated after program completion

Part VI

Strategic Recommendations for the Future

Building M&E Capacity

Part VI: Strategic Recommendations for the Future

Purpose of Part VI

This section synthesizes the **conceptual framework** from Parts I and II and the **practical insights** from parts III, IV, and V into recommendations for SFSD.

Creating M&E Capacity

Recommendation #1: Create an M&E position in SFSD

Rationale: In face of the complexity that is involved in carrying out sound evaluations as well as of the time and effort they require, we recommend that SFSD hires or assigns a current employee to be in charge of M&E-related tasks. The new position (M&E Manager or Coordinator) needs to be dedicated full-time to M&E and the selected individual would be the main person responsible for **a) deciding, from a technical perspective, which projects need to be evaluated**, **b) which type of evaluation to use on the selected projects, c) conducting the implementation of simple evaluation projects, such as process or outcome evaluations, d) helping operate the performance measurement system of SFSD** and, **e) being the point-of-contact with external organizations & consultants in issues that relate to M&E**. Moreover, M&E efforts need buy-in from the entire organization and should not be championed by the M&E Manager/Coordinator only. For that reason, the M&E Manager/Coordinator needs be able to mentor/guide co-workers by answering questions and providing on-the-job coaching whenever appropriate. Below is a list of necessary skills for the position:

Skills to Look for in an M&E Manager:
Good understanding of youth, decent work, and poverty issues
Understanding of quantitative and qualitative tools for data collection (surveys, focus groups, etc.)
Ability to check and clean data
Ability to analyze data
Report-writing and analytical skills
Facilitation, teaching, and reflective learning skills
Excellent communication and presentation skills
Proficiency in Microsoft Office and other statistical software as needed

Table retrieved from Taqeem Initiative's report *Making Learning Count*

While the specific tasks of an M&E Manager/Coordinator varies from organization to organization, we are including in this report a sample job description which was used by USAID as an example in its report titled "*Hiring M&E Staff – Guidelines and Tools for Locating and Hiring Strong Monitoring & Evaluation Candidates*"[1] (Hagens & Sharrock, 2008).

[1] Only the information that was considered to be relevant to SFSD was included in this report. For the full job description, please refer to the link on Appendix A: Useful Resources.

Sample Job Description – Catholic Relief Services, Malawi

Job Title: Monitoring & Evaluation Coordinator (I-LIFE Programme)

Purpose: The primary responsibility of this position is to provide programmatic support and technical assistance to I-LIFE through collaboration with partners to ensure program quality. M&E Coordinators will play a key role in liaising with partners on all programmatic aspects of I-LIFE. Key functions of this position will include monitoring and evaluation (M&E), learning, program implementation, and representation.

Key Job Tasks:

- Provide guidance and support to I-LIFE partners in the development and implementation of monitoring tools, which will include, but not be limited to, data collection, analysis and reporting on program indicators;
- Collaborate with I-LIFE partners in planning, designing and undertaking program evaluations/assessment;
- Foster learning within I-LIFE, through sharing lessons learnt, evaluations results, as well as playing a key role in Technical Working Groups among other ways;
- Ensure standardization of M&E tools and methodologies across implementing partners;
- Provide guidance and support to I-LIFE partner(s) in preparing and submitting timely quality progress reports;
- Provide leadership in the development and management of I-LIFE information management system;
- Contribute to development, growth and maintenance of I-LIFE resource center;
- Collaborate with I-LIFE partners to plan, conduct and/or facilitate programming related capacity-building activities, e.g., trainings.
- Act as key M&E liaison for assigned I-LIFE partners;
- Participate in field visits to partner working area to monitor progress and provide guidance;
- Participate in I-LIFE technical working groups and IGM meetings.

Required Qualifications:

- Minimum of a University degree in Social Science, Economics or Rural Development or in any related field. With extensive experience in monitoring, evaluation and learning;
- Sound experience in participatory research methodologies;
- Knowledge of word processing, spreadsheets, data entry and analysis packages and data bases like MS Word, Excel and SPSS);
- Demonstrate professional maturity and good interpersonal skills for teamwork, and good written and verbal communication skills;
- Ability to work to meet deadlines in multiple tasking environments;
- Previous experience working on Title II programming;
- Previous experience working in a consortium.

Desired

- Post-graduate certificate in a related field and experience with surveys, sampling methodology, data analysis and reporting;
- Sound report writing, interpersonal, facilitation, and communication skills.

Benefits: By having a professional dedicated full-time to M&E purposes, we believe that SFSD will be able to gradually incorporate M&E practices to its portfolio and eventually carry them out systematically. Without having a clear assignment of M&E responsibilities to an M&E professional, it can be challenging for organizations to implement evaluation projects in a non-sporadic fashion.

Recommendation #2: Establish a Performance Measurement System

Rationale: While SFSD currently keeps track of some information (activities & outcomes) of the programs it funds, we recommend that the foundation **creates a centralized database to curate key indicators from most or all of its programs**—thus constituting a performance measurement system.

Part IV of this report provided leads on what these key indicators could look like by bringing examples from the three case studies, but we would like to emphasize that the decision on which indicators to include is contingent on what aspects of the programs are important to highlight and what key message they would convey once reported. It is important to consider indicators beyond those that relate directly to the youth. A performance measurement system that has indicators on the **direct** (family, neighborhood, etc.), **local** (local economy and government), and **societal environment** (country demographics) in which the program participants are embedded is likely to yield a more refined assessment than otherwise. The Appendix A: Useful Resources has a list of examples of such indicators.

It is expected that the performance measurement system would be partially operated by the new M&E position as well as by the foundation's administrative staff, while under guidance of the M&E Manager/Coordinator's guidance.

Benefits: Having a solid performance measurement system helps transform the findings of individual program evaluations into lessons and improvements at the organizational level. Therefore, knowing how the organization is performing in terms of key indicators that range across most of its programs is a powerful piece of information for strategic decision-making. It also increases transparency in relation to community stakeholders (donors, clients, partner NGOs, etc.), further enhancing the organization's credibility.

Conducting Evaluations

Recommendation #3: Use a Humphrey School of Public Affairs "field experience team" to carry out a process/outcome evaluation of the Vocational Productivity Program

Rationale: One of the deliverables of this project includes a draft evaluation plan for the "Vocational Productivity Program", which highly suitable to a **process and outcome evaluation**. Therefore, SFSD can make use of the Humphrey School of Public Affairs (HHH) field experience teams in order to implement such evaluations at a low cost.

The Master of Development Practice (MDP) program at the HHH requires students to complete a summer, 8-10 week international field experience as part of their degree. SFSD could answer to HHH's request for

proposal and ask a student team to carry out the process and outcome evaluation according to the draft plan provided in Part V[2].

The HHH request for proposal is typically issued at the end of every year and field experience teams carry out their projects in the period between May and August. The student teams are required to visit the site, and thus, they are able to carry out the evaluation on-site. For further information, SFSD's staff may contact the MDP Program Coordinator, Dr. David Wilsey:

Dave Wilsey, PhD
Humphrey School of Public Affairs (HHH) | University of Minnesota
301 19th Avenue South, Room 258
Minneapolis, MN – 55455
+1 612-625-7062
dwilsey@umn.edu

Benefits: The advantage with the student teams is that SFSD can utilize their services at no upfront cost and many, if not all, students have significant background in M&E as part of their graduate studies at HHH. Field experience teams also need to be on the ground—a core requirement of any evaluation. This, combined with the Vocational Productivity Program's suitability for a process and/or outcome evaluation makes it an ideal opportunity for this continued partnership.

Additional Considerations: The obvious question in this regard is why SFSD should consider the services of external versus internal evaluators (if it decides to set up an M&E department). While there is a growing trend of internal evaluations worldwide (Love, 2005; Sonnichsen), the services of external evaluators can offer the follow *advantages*:

- External evaluators can view the program from a different perspective;
- Program participants and staff may be more honest with an external evaluator;
- They have greater independence in making recommendations;
- They can bring in a higher degree of objectivity in determining the effects of the program.

However, external evaluators (or evaluations) also pose some *disadvantages* in terms of limited knowledge about the organization and its culture or relying on cooperation of internal members to gain access to data. They also tend to be more expensive (for instance, for involving travel arrangements).

Recommendation #4: Partner with specialized organizations when an impact evaluation is needed

Rationale: Conducting an impact evaluation requires specialized skills and knowledge in quantitative data collection and analysis (as per Part II of this report). For this reason, we recommend that **SFSD partners with reputed organizations that conduct rigorous impact evaluations whenever it wishes to carry out an impact evaluation of one (or more) of its programs**. The following organizations can serve as potential partners for SFSD:

[2] It is expected that the draft plan would undergo changes at discretion of the field experience team. Evaluation plans are seldom designed at a distance and some adjustments are expected from a team of consultants that is on the field.

a. Taqeem Initiative (International Labour Organization)

Taqeem is an initiative of the ILO in the Middle East/North Africa region that works to assist organizations with youth employment and entrepreneurship intervention in enhancing their ability to evaluate their program[3]. It does this through what it calls a "Community of Practice" (CoP) approach. Agencies that are part of this CoP are competitively selected by the Taqeem Initiative to build capacity in monitoring and evaluation including impact evaluations. The funding for financial and technical assistance comes from the ILO *Fund for Evaluation in Youth Employment*, which is in turn supported by agencies such as the International Development Research Centre (IDRC), the International Fund for Agricultural Development (IFAD), the International Initiative for Impact Evaluation (3ie), Silatech and the Swiss State Secretariat for Economic Affairs (SECO).

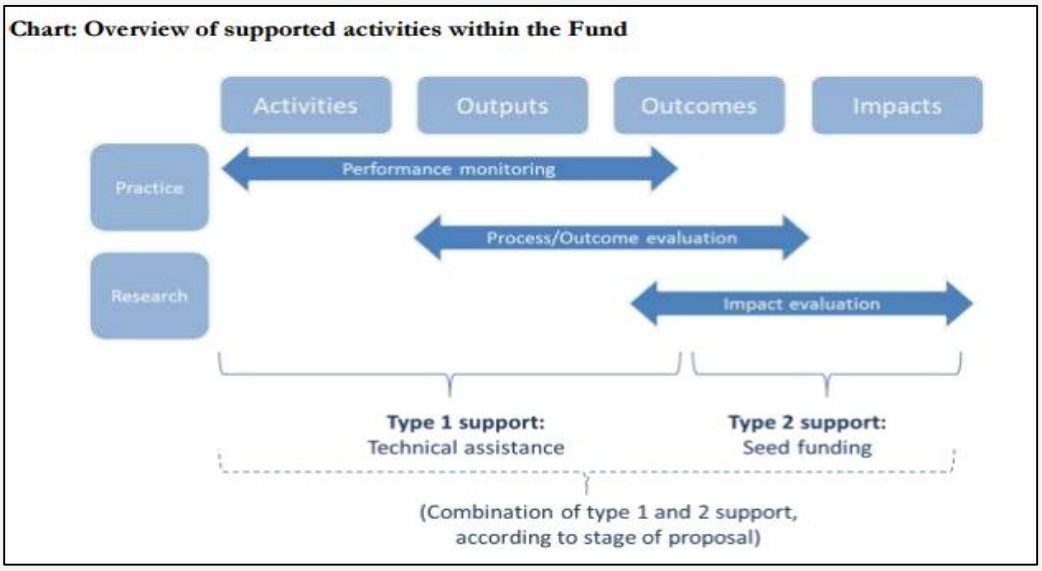
The Fund targets two kinds of organizations:

- Implementing organizations, which include public, non-governmental or private organizations with operations in one of the MENA countries;
- Research institutions: based worldwide, but studying effects of employment and enterprise development interventions implemented in the MENA region.

Broadly, the capacity building assistance through Taqeem involves the following components:

- Technical assistance: It involves evaluation clinics which provide basic M&E training and last for 3-4 days as well as guidance from experts who help organizations develop and revise M&E plans including impact evaluation plans;
- Financial assistance: Organizations may get small grants to cover some parts of implementing the evaluation plan. It also involves seed funding of up to a total of $650,000 dollars for selected impact evaluations.

The following graphic distinguished the two types of assistance provided under the Fund.



[3] Excerpt retrieved from http://www.ilo.org/employment/areas/youth-employment/WCMS_458221/lang--en/index.htm

Additional resources provided through the Initiative include the advanced *Executive Evaluation Courses* on youth labor market programs and the *Fund for Evaluation in Youth Employment online platform* that provides users resources such as data collection protocols and evaluation reports.

b. Abdul Latif Jameel Poverty Action Lab (J-PAL)

The Abdul Latif Jameel Poverty Action Lab (J-PAL) is a network of 131 affiliated professors from over 40 universities who work to provide rigorous scientific evidence that informs policy around poverty and development issues. It was founded in 2003 at the Department of Economics in Massachusetts Institute of Technology by professors Abhijit Banerjee, Esther Duflo, and Sendhil Mullainathan. They partner with implementing organizations such as governments and NGOs, donors including foundations that fund evaluations and research centers that run randomized evaluations worldwide.

JPAL's Youth Initiative, funded by the Nike Foundation, specifically focuses on evaluating interventions that "help disadvantaged youth to successfully navigate the transition from adolescence to adulthood."

Their work in the area of evaluation of youth employment interventions includes projects such as:

1. *Evaluating the Effects of Entrepreneurship Edutainment*, Egypt: in which researchers partnered with Bamyan Media to measure the impact of a television show, online entrepreneurial support activities, and networking events on youths' attitude towards business creation, business practices, business skills and employment status of the viewers.

2. *Returns to Apprenticeship Training*, Ghana: where researchers are involved in assessing the impact of a government apprenticeship program that pairs young people who have limited education with master trainers operating small businesses for a one-year training period. In addition to examining the targeting and recruitment process, benefits for apprentices and trainers, the researchers will also examine the cost-effectiveness of the program.

3. *The Impact of Entrepreneurship Training for Women*, Uganda: where researchers evaluated the effect of standardized business skills training delivered through a training program called Women Mean Business on business performance of women in four cities in central Uganda.

Additionally JPAL runs a number of executive education courses for researchers, policy-makers and donors. It offers four types of courses: Evaluating Social Programs, Evaluating Social Programs (Advanced), M&E Management, and Measurement. Further, it works with governments, NGOs, and donors to scale up programs found to be effective with a view to promote evidence-informed policy.

c. International Initiative for Impact Evaluation (3ie)

3ie is a grant making NGO that funds and produces high-quality evidence to promote evidence-informed policies in international development. Since its founding in 2008, 3ie has awarded over 200 grants (146 impact evaluations, 33 systematic reviews and 38 other studies) in over 50 countries.

The agency's work in impact evaluation involves:

- Maintaining an impact evaluation database which has information on over 2,500 impact evaluations. This includes information on over a 100 impact evaluations funded by 3ie covering sectors such as agriculture and rural development, health and nutrition, and disaster management among others;

- Maintaining an expert roster which allows policy makers and practitioners to identify experts for designing and conducting impact evaluations;
- Replication program, which is a form of research validation in order to lend or test the credibility of impact evaluation findings. Valid findings in replication studies enhance the credibility of the impact evaluation whereas invalid findings cautions policy-makers regarding using the evidence;
- RIDIE, which is a prospective registry of impact evaluations related to development in low- and middle-income countries;
- A collection of reference materials, tool-kits, downloadable documents and other resources related to methodology and best practice in impact evaluation;
- Professional services, such as design clinics and impact evaluation external quality control.

Recommendation #5: Integrate evaluation into program design

Rationale: As covered by Parts I and II of this report, information collection can be a major cost driver of many evaluations, particularly in the case of retrospective evaluations (that take place after the program is complete). Therefore, **we recommend that, to the extent possible, SFSD embeds M&E practices into the design of the programs it funds**. A common limiting factor to many program evaluations is having to collect information that was available when the program started but is no longer at completion. For instance, it is much easier (and more accurate) to test the participants' job-searching skills through an entry test than having to ask each participant, at the end of the program, what were their job-searching skills before the program started. This hypothetical situation can be common when evaluations are not planned for when the program is being designed. Therefore, the M&E Manager/Coordinator needs to participate in the design of new programs so that M&E concerns (data collection steps, program-specific indicators, type of evaluation, estimation of M&E-related expenses, reporting timeline, etc.) are addressed *prior* to the program's implementation.

It is also worth noting that even well-structured evaluations do not make up for inadequate program design. The success of a program is ultimately contingent on its activities, and integrating M&E to program design only ensures that the program's results can be verified upon completion, but it does not influence whether or not the program is successful.

Benefits: When evaluations are planned for in advance the cost of collecting information can be significantly reduced, keeping track of program participants becomes easier, and the *quality* of the findings of the evaluation are greatly enhanced. Moreover, depending on the program design, some impact evaluations are entirely *not viable* when not planned for in advance.

Post-Evaluation

Recommendation #6: Share the knowledge when robust evaluations are carried out

Rationale: Besides improving the effectiveness of its own programs through evaluation, SFSD could also help diminish the evidence gap that exists when it comes to youth employment training programs, particularly in MENA. Some international organizations collect and curate evaluations from around the world as a means to share information and help like-minded organizations advance their mission. For instance, the Youth Employment Inventory (a World Bank-initiated partnership among many international organizations) compiles evaluations studies of worldwide youth employment interventions, and makes their findings available online. Sending complete, robust evaluations their way is a form of helping the

spread the knowledge as well as further validating the results of SFSD's projects. **Thus we recommend that, whenever robust evaluations are conducted, SFSD shares the results.** Below is a list of possible dissemination outlets provided by the *"Measuring Success of Youth Livelihood Interventions"* report.

Youth Employment Inventory

Evaluations are submitted to YEI through the link: http://www.youth-employment-inventory.org/inventory/submit/. SFSD will be asked to provide information on a) *program activities* (type of intervention), b) *status of the intervention* (completed, ongoing), c) *participants' characteristics* (youth at risk, male/female, low income, etc.), d) *evaluation type* (monitoring, process, impact evaluation, etc.), and e) *costs of program & evaluation*. The submitted information is then reviewed by YEI, and once approved, becomes available to any visitor to the website.

Naturally, the YEI makes available the same type of information from other projects—making it a good resource for SFSD to verify the evidence of success (or failure) of certain types of intervention, on a case-by-case basis.

Youth Employment Network Website

The Youth Employment Network website (http://yenclinic.groupsite.com/main/summary) "*connects policy makers and the practitioners' community around the topic of youth employment and evaluation*". The network accepts submissions of **impact evaluations** only and also provides opportunities for connecting with professionals from other organizations that support youth employment initiatives.

Eldis.org

Eldis.org is a knowledge-sharing platform that is not constrained to M&E—it compiles country profile reports, humanitarian situation assessments, labor statistics, and other documents from credible organizations around the world. However, they accept submissions of **impact evaluations** and host documents on their platform once they are reviewed and approved.
(http://www.eldis.org/go/contribute#.V1tX-7srLtQ)

Other Organizations and Platforms

Besides the online platforms listed above, SFSD may choose to present the findings of robust evaluations to local partners, government authorities, inter-governmental organizations, and others foundations via conferences or face-to-face dissemination. As is with SFSD, there are many other actors looking for evidence regarding youth employment programs who could make use of the evidence created by a successful M&E project.

Benefits: Sharing evidence of youth employment interventions' success/failure, particularly from the "evidence-starved" MENA region, is of great value for other organizations and foundations like SFSD. It helps with establishing best practices and consolidating successful program designs, thus contributing to the common good. In return, it can help SFSD grow even further its global footprint among similar organizations.

Building M&E Capacity

Bibliography

Main Sources

Hempel, K., & Fiala, N. (2011). *Measuring Success of Youth Livelihood Interventions: A Practical Guide to Monitoring and Evaluation (Rep.)*. Retrieved June 8, 2016, from Global Partnership for Youth Employment website: http://www.iyfnet.org/sites/default/files/gpye-m&e-report.pdf

Making Learning Count: Effective Monitoring and Evaluation of Youth Employment Programs in the Arab World (Rep.). (2015, June 2). Retrieved http://www.ilo.org/wcmsp5/groups/public/---ed_emp/documents/publication/wcms_372803.pdf

Russ-Eft, D., & Preskill, H. (2009). *Evaluation in Organizations: A Systematic Approach to Enhancing Learning, Performance, and Change* (2nd ed.). Basic Books.

Complementary Sources

Betcherman, G., Godfrey, M., Puerto, S., Rother, F., & Stavreska, A. (2007). *Global Inventory of Interventions to Support Young Workers Synthesis Report.* Retrieved from http://siteresources.worldbank.org/EXTECAREGTOPYOUTH/Resources/Youth_Employment_Inventory.pdf

Gertler, P., Martinez, S., Premand, P., Rawlings, L., & Vermeersch, C. (2011). *Impact Evaluation in Practice.* Retrieved from http://siteresources.worldbank.org/EXTHDOFFICE/Resources/5485726-1295455628620/Impact_Evaluation_in_Practice.pdf

Hagens, C., & Sharrock, G. (2008). *Hiring M&E Staff: Guidelines and Tools for Locating and Hiring Strong Monitoring & Evaluation Candidates*. Retrieved from http://static1.1.sqspcdn.com/static/f/752898/9984725/1296501643587/MEmodule_hiring.pdf?token=xp2ZYF518QZQZuhfOPttysPAJ80%3D

International Labour Organization. (2012). ILO Policy Guidelines for Results Based Evaluation. Retrieved from: http://www.ilo.org/wcmsp5/groups/public/---ed_mas/---eval/documents/publication/wcms_168289.pdf

Krueger, R. (n.d.). *Outcome Evaluation*. Retrieved from http://www.tc.umn.edu/~rkrueger/evaluation_oe.html

Miller, E., & MacGillivray, L. (2002, April). *Youth Offender Demonstration Project Process Evaluation* (Final Report) (Rep.). Retrieved from http://files.eric.ed.gov/fulltext/ED470717.pdf (ERIC Document Reproduction Service)

Sabatelli, R., & Anderson, S. (2005, September). *Assessing Outcomes in Child and Youth Programs: A Practical Handbook* (Rep.). Retrieved from http://www.uwex.edu/ces/4h/evaluation/documents/ChildYouthOutcomeHandbook2005.pdf

Saunders, R., Evans, M., & Joshi, P. (2005, April). *Developing a Process-Evaluation Plan for Assessing*

Health Promotion Program Implementation: A How-To Guide. Health Promotion Practice, 6(2), 134-147. doi:10.1177/1524839904273387

Youth Empowerment Program Evaluation Report (Rep.). (2010, March). Retrieved from http://www.iyfnet.org/sites/default/files/YEP_EvalReport_Kenya_(NAIROBITS).pdf

Appendix A

List of Useful Resources

Building M&E Capacity

Appendix A: List of Useful Resources

Capacity Building Resources

My M&E
Interactive web platform with training sessions, toolkits, and other resources on M&E.
Link: http://www.evalpartners.org/

ILO International Centre Course on M&E for Youth Employment Projects
ILO's training institute which provides courses on varied topics related to Youth Employment.
Link: http://www.itcilo.org/en

International Program for Development Evaluation Training (IPDET)
A World Bank-sponsored training program that provides courses on M&E for NGOs, private corporations, research institutes, and others.
Link: http://www.ipdet.org/

J-PAL Course on Impact Evaluation
World-renowned J-PAL provides courses and workshops to prepare development practitioners for Impact Evaluations projects.
Link: https://www.povertyactionlab.org/training-0

World Bank Impact Evaluation Workshops
The Spanish Impact Evaluation Fund sponsored many workshops up until 2009. The materials from such workshops are available online.
Link: World Bank Workshops

YEN Evaluation Clinics
The Fund for Evaluation in Youth Employment sponsors workshops and brings together the community of researchers and practitioners around the topic of youth employment evaluation.
Link: http://yenclinic.groupsite.com/main/summary

Databases of Impact Evaluations

International Initiative for Impact Evaluations (3ie)
3ie's webpage contains selected Evidence Gap Maps, Policy Briefs, and Systematic Reviews of impact evaluations that may be of use for SFSD's staff.
Link: http://www.3ieimpact.org/en/evidence/

Abdul Latif Jameel Poverty Action Lab (J-PAL)
J-PAL makes the impact evaluations conducted by their affiliates available on their website.
Link: https://www.povertyactionlab.org/evaluations

Innovations for Poverty Action (IPA)
Database of Impact Evaluations searchable by topics, program areas, and country.

Link: http://www.poverty-action.org/search-studies

Youth Employment Inventory (YEI)

Database of several types of evaluations, not restricted to only Impact Evaluations. Database includes youth employment programs only.
Link: http://www.youth-employment-inventory.org/

World Bank Development Impact Evaluation (DIME) Initiative

List of Impact Evaluations sponsored by the DIME initiative.
Link: Dime Initiative

Other Resources

ILO I-track Database

This database stores evaluation-related documents, such as evaluation planning workflows, recommendations, lessons learned, etc. It also provides an "Evaluation Consultant" list. Follow the 'generic log-in instructions' in order to get access to the database.
Link: http://www.ilo.org/evalinfo/

ILO Good Practices on Youth Employment

A website that provides good practices on youth employment programs. Searchable by country and thematic area.
Link: http://youthpractices.org/index.php?lang=en

YouthPol eAnalysis

An ILO's database of youth employment policies around the world.
Link: http://www.ilo.org/dyn/youthpol/en/f?p=30850:1001:0::NO:::

Silatech

Silatech's website provides reports on youth employment issues in MENA. It is also a partner of the Taqeem Initiative.
Link: http://silatech.com/