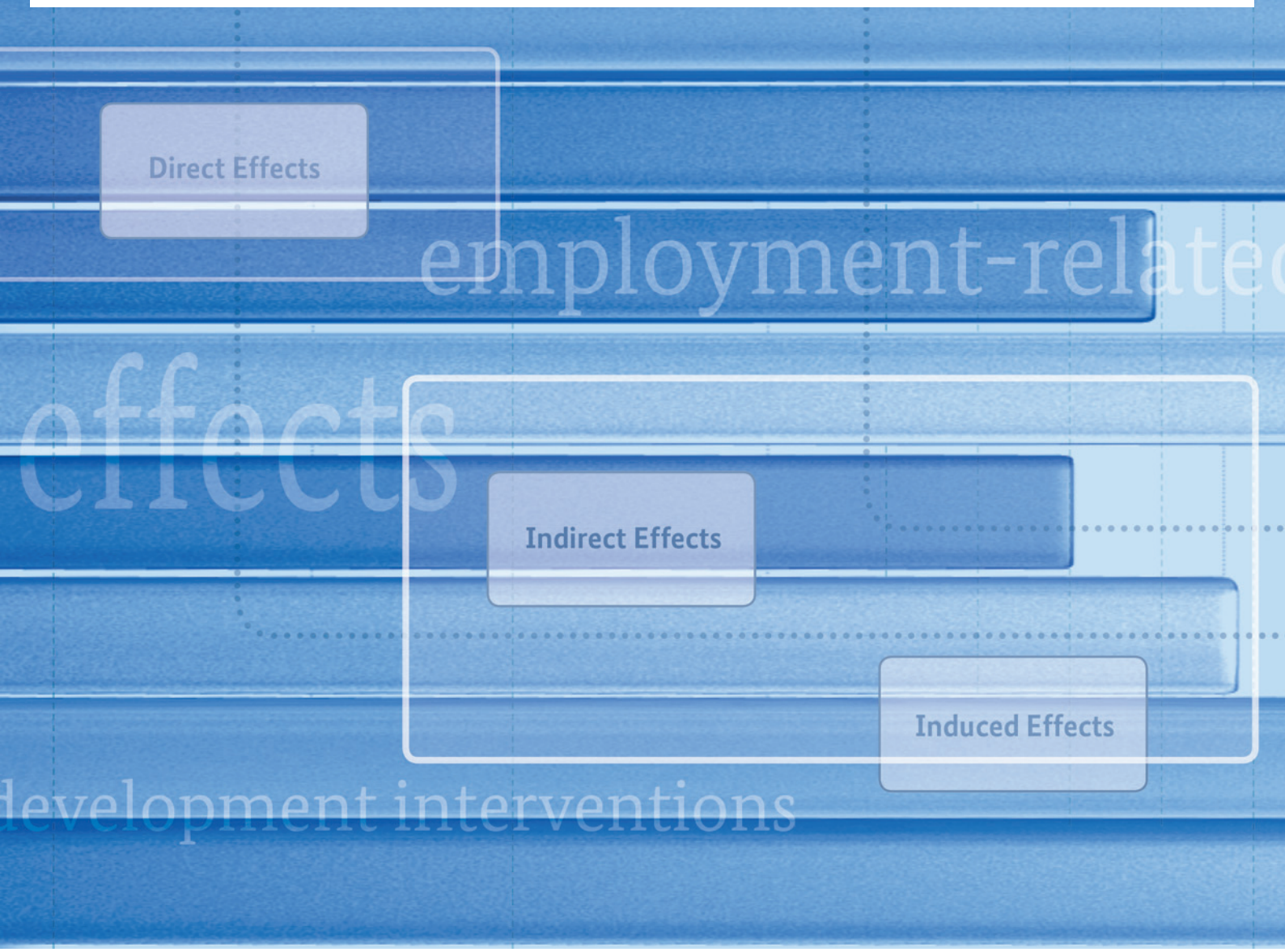
Federal Ministry
for Economic Cooperation
and Development



A Systematic Framework for Measuring Employment Impacts

of Development Cooperation Interventions

Published by giz Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH

A Systematic Framework for Measuring Aggregate Employment Impacts of Development Cooperation[1]

Jochen Kluve
Humboldt-Universität zu Berlin and
Rheinisch-Westfälisches Institut für Wirtschaftsforschung

Jonathan Stöterau
Rheinisch-Westfälisches Institut für Wirtschaftsforschung

The analysis, results and recommendations in this paper represent the opinion of the author(s) and are not necessarily representative of the position of the Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH.

1 We gratefully acknowledge many helpful discussions and comments from colleagues at GIZ and KfW in Germany and Morocco, from conference participants at the 2014 PEGNet conference in Lusaka, from participants of a GIZ workshop in Cairo and of BMZ workshops in Berlin and Bonn.

Contents

Abstract.

This paper summarizes the key features of a systematic framework for assessing the aggregate employment impacts of a portfolio of development cooperation interventions. The methodological approach follows a bottom-up procedure based on three steps: (i) estimating employment effects at the intervention level, net of the counterfactual scenario; (ii) estimating (economy-wide) employment impacts of interventions taking into account employment-related indirect effects of the intervention such as displacement, substitution and multipliers; (iii) aggregating these employment impacts across the portfolio and deriving comparable parameter values for employment effects. We discuss these steps, along with two preparatory steps that enable an identification of projects for which a detailed evaluation may be most relevant and feasible. To this end, we develop a classification of intervention types from an employment perspective and propose an approach to judge their evaluability ex-ante. Finally, we discuss how these bottom-up estimates can feed into a system of institutional learning about employment impacts, based on the specification of an indicator for employment outcomes that can be compared and aggregate across heterogeneous development projects. The paper is based on an exploratory study conducted for German development cooperation; hence there is a necessary focus on a framework applicable in this specific context.

Jochen Kluve
RWI and Humboldt-Universität zu Berlin
Invalidenstr. 112
10115 Berlin
jochen.kluve@rwi-essen.de

Jonathan Stöterau
RWI
Invalidenstr. 112
10115 Berlin
jonathan.stoeterau@rwi-essen.de

1. Introduction

Employment and job creation are increasingly becoming a key focus in development cooperation. This is substantiated, for instance, by the fact that the World Bank's 2013 World Development Report focused exclusively on "Jobs". The report comes to the conclusion that jobs serve to "boost living standards, raise productivity and foster social cohesion". Against the fact that employment has been identified as a key driver of poverty reduction (Fields 2013), many governments and donors are keen to identify effective policies and programs that can help create and sustain employment opportunities.

The importance of jobs for poverty reduction is a critical argument for development cooperation to measure and increase their impacts on employment in partner countries. But the complexity of the topic makes measurement challenging and potentially error-prone (Fowler and Markel, 2014). Employment comes in several dimensions. Frequently not only the creation of new jobs matters for development. Other aspects – such as the quality or sustainability of jobs – are sometimes of even greater importance.[2] Moreover, many development programs may have large (and foremost) indirect or induced effects on employment outcomes – some of which may be unintended or negative. As a consequence, relatively little systematic knowledge and guidance is available about the employment impacts of development cooperation programs.

Against this background, the German Federal Ministry for Economic Cooperation and Development (BMZ) commissioned a study with a three-fold task. First, to test the feasibility of assessing aggregate employment effects of all German development cooperation programs (incl. technical and financial assistance) within a short-term portfolio assessment in one pilot country. Second, to develop evaluation designs to estimate net employment impacts based on the available data from projects' Monitoring and Evaluation (M&E) Systems and/or secondary data available in the country. Third, to elaborate on ideas for assessing potential employment effects of interventions prior to program implementation (ex-ante). Morocco was chosen as a pilot country since the region of Northern Africa and Middle East is characterized by a large share of youths in the labor market and high population growth, i.e. a context in which effective employment creating policies are of particular relevance.

A detailed analysis of the portfolio of German development interventions in Morocco undertaken in the resulting study (RWI 2013) suggests that a large share of programs aim to have direct or indirect employment effects. But an assessment of available primary and secondary data showed that no reliable estimate of aggregate employment creation of German development cooperation in Morocco could be derived in the short-term. This result likely does not pertain to the specificities of Morocco and/or German development cooperation programs. As a consequence, the research team was asked to explore methods that would allow an estimation of aggregate job-creation as part of portfolio assessments in the future.

Against this background, the primary objective of the research was to develop a methodological framework that allows estimating economy-wide employment impacts ("policy-relevant") across heterogeneous intervention types ("universal") under prevailing constraints ("practical"). This paper summarizes the findings of that study (RWI 2013) and the key characteristics of the methodological framework suggested therein. The aim is to provide an input to an ongoing discussion about ways to improve the evidence regarding development cooperation impacts on employment. Given the background of the project, it is evident that the framework developed in this paper necessarily focuses on an approach applicable in the context of German development cooperation. No claims regarding a broader applicability are made.

We describe the features of a systematic bottom-up procedure to arrive at an aggregate impact estimate for a given portfolio. To present this concept, we explicitly distinguish between *gross* and *net* employment *effects* of a specific intervention and (economy- or market-wide) employment *impacts* of development projects. The proposed methodology consists of three steps:

2 The ILO estimates that 60% of the labor force in developing countries works in the informal sector, with 34% earning below $2 per day. Nearly half face vulnerable job conditions (ILO 2014).

i. In the first step, employment effects of a specific development intervention are estimated. The goal of this step should be to assess *net effects for the target population* (i.e. accounting for the counterfactual outcome), which typically requires an adequate impact evaluation design. Since, in practice, data and resource constraints often limit the possibilities to implement rigorous evaluation designs, we discuss a second-best option – based on converting gross employment effects via a parameter for the counterfactual. However, the second-best option should only be regarded as a methodological "last resort" in a case in which there is no possibility at all to implement a rigorous research design, but plausible values for employment effects are desired for aggregation purposes.

ii. The second step proceeds to estimate the *economy-wide employment impacts* of the intervention, taking into account intended or unintended indirect effects (e.g. substitution, displacement and multiplier effects). As in the first step, we present a second-best option to derive plausible parameters for these effects from secondary sources or previous program evaluations, since a precise estimation of these indirect effects may not be feasible in every given context. Whereas in the first step the second-best option is an absolute exception, in the second step the second-best option will often be the only feasible way to proceed.

iii. Finally, the third step aggregates impacts across all interventions in the portfolio to provide an estimate for the overall, *aggregate employment impact* in the economy under study. Moreover, at this stage, evaluation results and parameter estimates are integrated into a "learning system" at the institutional level. On this basis, parameter estimates can inform the ex-ante and ex-post assessment of employment impacts in related development programs.

The stepwise procedure implies a "bottom-up" approach to estimate aggregate economy-wide employment impacts of a portfolio of development interventions. An alternative methodological approach would be top-down or semi-top-down approaches, i.e. estimating changes in aggregate indicators on the basis of input-output tables for the entire economy or specific sectors.

However, the bottom-up approach has several advantages in the context of informing the work of development agencies (for a more detailed discussion see RWI 2013): first, it provides a possibility to assess *economy-wide employment impacts* of development programs within regular evaluation frameworks (which, we argue, is the relevant measure for any program evaluation interested in welfare impacts). Second, the key ingredient, a rigorous estimation of employment effects at the intervention level is well understood, and a toolbox of appropriate evaluation methods exists. Third, the determination of aggregate effects works under comparatively few assumptions, which can be clearly presented for each step of the bottom-up procedure. The procedure thus delivers a transparent picture regarding the precision and reliability of aggregate figures. Finally, the bottom-up approach identifies how each intervention and each project contributes to the overall employment impact of a portfolio. Hence, the proposed procedure delivers the basis for a comparison of the effectiveness of different interventions and development projects across sectors and countries.

Additionally, we introduce a procedure to systematically identify which interventions of a portfolio are most likely to impact on employment, and are thus most relevant to include in the application. To this end, we develop a typology of development interventions that specifically takes into account the intervention logic regarding employment effects. This assessment can be an (optional) preparatory step in the analysis of large portfolios, in the case in which no a-priori considerations determine the selection. The proposed classification may also provide the basis for an identification of programs for which an implementation of rigorous evaluations into program design will be particularly beneficial.

Finally, we describe how the method could deliver the evidence for a systemized learning about employment effects of development cooperation projects in the long run. Within the proposed framework, each application of the method produces results that inform the next applica-

tion, thus sequentially improving the impact estimates. If implemented purposefully and consistently, this systematic bottom-up procedure can reduce the time and costs associated with assessing employment effects for development projects ex-post and ex-ante.

The proposed framework is practical, in the sense that it allows for different levels of methodological rigor at each step, depending on the available data and resources. Specifically, we describe alternative strategies (second-best options) to derive intervention-level employment effects also in situations where time constraints, data limitations or financial reasons do not allow for full-scale rigorous impact evaluation. By way of this practical approach, the framework becomes universal, as it allows to arrive at a transparent and plausible estimate for intervention types which are otherwise difficult to evaluate. This allows continuing with the portfolio analysis, based on transparent and evidence-based estimates for such interventions. This feature is enhanced through the "learning system", where each evaluation of employment impacts can reduce the resources needed to arrive at a plausible estimate for similar interventions.

Notwithstanding the idea to design a practical and universally applicable framework, two aspects are to be kept in mind: first, the rigorous estimation of net employment effects via an (early) integration of evaluations into program monitoring systems constitutes the only basis for any precise assessment of employment impacts. The proposed second-best options can only present a substitute to allow the realization of an ex-post portfolio analysis. Second, the assessment of aggregate net employment impacts for a portfolio is generally an immensely complex endeavor. Even a systematic, clearly structured and practicable framework cannot be trivial in its details. This rather short paper can only present a rough conceptual overview but does not attempt to provide all details of its implementation.

The remainder of the paper presents the main elements of the framework along the proposed bottom-up procedure: The following section clarifies the terminology used in the remainder of the paper and introduces the two optional steps to arrive at an informed project selection. Sections 3 and 4 discuss how employment effects and impacts, respectively, can be estimated within the bottom-up procedure. Within each section, we first present the first-best option and then sketch out alternative (second-best) solutions for estimating employment-relevant parameters. The final section 5 discusses the requirements which are needed in order to aggregate and compare estimates from different evaluations. To this end, it formulates an indicator for employment effects which can be aggregated and compared across heterogeneous development interventions. The section finishes by discussing how the approach can be integrated into a system of institutional learning.

Back to Contents

2. Preparing the analysis: terminology and project selection

The proposed bottom-up approach consists of three steps: (i) estimating employment *effects* at the intervention level, net of the counterfactual outcome; (ii) estimating (economy-wide) employment *impacts* of interventions, comprising employment-relevant indirect and induced effects such as displacement, substitution and multipliers; (iii) *aggregating* these employment impacts across the given portfolio. These steps are presented in more detail in the subsequent subsections.

To present a concise reading of this framework, section 2.1 aims at clarifying the terminology of development programs regarding employment impacts. While the terminology of key components of program implementation is fairly well established, a variety of definitions for the various types and levels of employment effects are found in the economic literature.[3] For instance, the concepts of "net effects" and "impacts" are often used interchangeably in the literature. Moreover, no unified conceptual approach exists to differentiate between indirect/induced effects, externalities/spillovers or second-/third-order effects. Understanding and distinguishing between them is important to inform the design of M&E-Systems, since the appropriate approaches to measurement differ for each one.

The second part of this section outlines an optional, preparatory step for the analysis of a portfolio regarding aggregate employment impacts: A systematic and transparent procedure to arrive at an ex-ante identification of interventions for which an in-depth analysis of employment impacts may be particularly valuable. The proposed approach is based on a classification regarding each program's potential *relevance* in terms of employment (section 2.2) and the *evaluability* of envisioned employment effects (section 2.3). To this end, we first outline a classification of programs based on the employment-specific results chain for various types of interventions. Based on this classification we then describe a possible approach to (ex-ante) assess their evaluability, even in the case in which no detailed knowledge about every program detail is available.

3 See also the discussions in DCED (2010), GTZ (2008), OECD-DAC (2002) and Kluve (2011).

2.1 Terminology

The basic terminology used to describe different levels of portfolios of development cooperation activities has been well established through the OECD DAC framework: generally, a portfolio (in a country or sector) comprises various *programs* which each involve one or multiple *interventions* with specific *inputs*, *outputs* and *outcomes*.[4] Moreover, each output of an intervention causes a set of *effects on individuals or firms*. In the remainder of the paper we distinguish explicitly between employment outcomes/effects and other (non-employment) outcomes/effects. Employment outcomes refer to, for example, the individual employment situation or the number of employees at the firm-level (see Section 5 for a discussion) at some point in time. It is also helpful to recall that an effect refers to the change in the level of an outcome variable before and after the intervention takes place (i.e. outcomes are levels, effects are changes).

Typically, each intervention has a *target group*, for which it aims to improve the level of specific outcomes. Our specific interest is in the target group for which employment outcomes are meant to be improved. We refer to *beneficiaries* as the subset of this target group that is actually affected by the outputs of the intervention. To provide one example: a training program may target young women in a specific region (but cannot provide training for all); the beneficiaries are those young women who participate in the training.

We distinguish specifically between *direct, indirect and induced employment effects*. This differentiation is essential in order to outline potential methodological approaches to estimate these effects. The specific terminology established in this paper aims to reflect the possible methodological designs and data collection efforts required to analyze these effects within M&E-Systems.[5] Box 1 summarizes the main terminology as it is used in this paper.

4 For more detailed definitions see OECD (2002).

5 The systematization of employment effects presented here is an extension and refinement to those used in the original study (RWI 2013), based on valuable feedback from several stakeholders of German development cooperation.

We refer to *direct effects* as all changes in outcomes (employment and other, non-employment) that are caused by the outputs of the intervention among its beneficiaries (Example: training causes an improvement in job-finding rates among training participants). Note that the subjects (individuals, firms, institutions) that are directly affected by the outputs are not necessarily the primary target group of the program in every instance. This is particularly the case for institutional development programs and political advisory programs. Although some employment effects may be achieved, these are generally negligible and not a primary goal of the intervention (e.g. training individuals at institutions as part of an institutional development program). Direct effects are usually considered a primary goal of the intervention, and should hence be part of any M&E System. However, direct employment effects are often mediated by changes in other outcome variables (Example: the effect of an intervention on firm-level employment is often mediated by productivity and/or sales). Only in specific instances would development interventions have an immediate direct employment effect (Example: infrastructure projects financed by development cooperation). It is therefore important to emphasize that we consider all effects on employment outcomes of beneficiaries as direct employment effects, whether mediated or not.

Indirect effects are changes in outcomes among individuals and firms that are part of the target group of the intervention but which are not directly affected by the outputs of the intervention. In the example from above, these may be other young women in the region, who do not participate in the training. Indirect effects are triggered by direct effects on the beneficiaries of the intervention. Similar to the above case of direct effects, also indirect employment effects are often mediated by changes in other outcome variables (Example: Increase in sales among beneficiaries of an intervention may reduce sales by other firms among the target group and (mediately reduce employment).

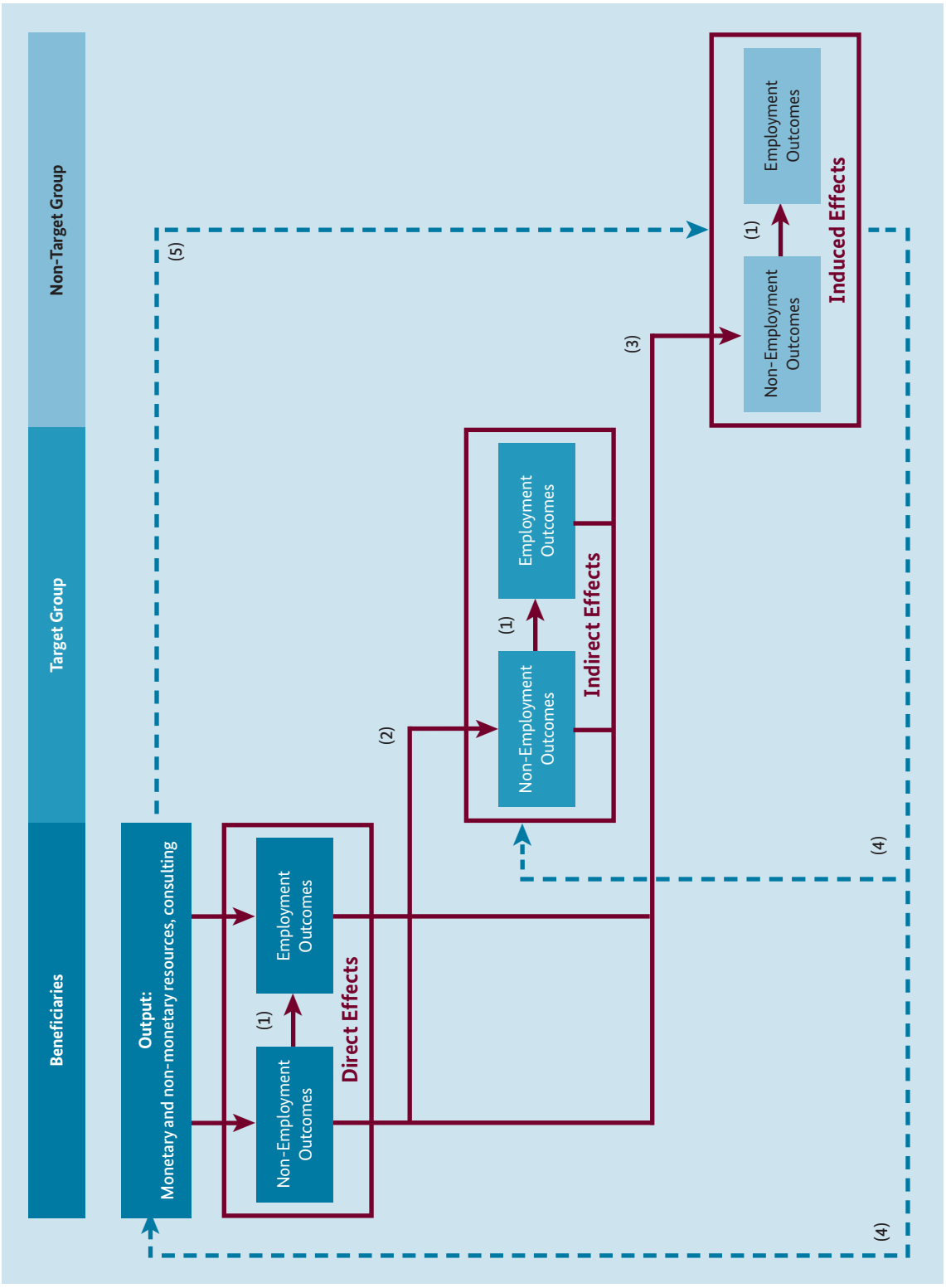
Finally, we refer to *induced effects* as all changes in outcome variables among individuals and firms that are not part of the target group of the intervention. To stick to the above example of training young women, the non-target group may be older women or men in the intervention region. The non-target group may also be young women outside the intervention area.[6] Similar to indirect effects, they are triggered by the direct effects of an intervention – i.e. they are also a result of the outputs of an intervention. In addition, they may also be brought about by the intervention's indirect effects. Induced employment effects are generally not a goal of the intervention; they can be acknowledged as an (intended or unintended) consequence.

This conceptual approach is presented in Figure 1 below. As an example, consider an institutional development program targeting SMEs in a given region: the outputs (e.g. training employees of institutions) cause an improvement in the local economic environment (direct non-employment effect), which in turn affects SME productivity and finally employment among firms (indirect employment effect). This, in turn, may have a positive (spillover/multiplier) or negative (displacement) effect on employment in firms outside the region (induced employment effect). The example is illustrated in Figure A1 in the appendix, along with two further examples illustrating this terminology for the cases of a training program (Figure A2) and an infrastructure project (Figure A3).

The above example also highlights the fact that, depending on the intervention and labor market characteristics, the indirect and induced effects may significantly increase or decrease the direct employment effect estimated only among the beneficiaries of an intervention. Moreover, the distinction between indirect and induced effects also provides a possibility to clearly differentiate between substitution and displacement effects – for which no unequivocal definition exists in the literature so far: we refer to substitution effects as employment effects on the target population that go at the expense of other subjects *within* the target population. In contrast, displacement refers to negative employment effects on subjects *outside* the target population as a consequence of direct and indirect effects. Hence, substitution is considered an indirect effect and displacement an induced effect. At the same time, multiplier effects (supplier- and consumption effects) can be both indirect

6 Note that, if an intervention generally aims at the improvement of individual-level employment outcomes (e.g. a training program), then the effects on firms (for example, through improved labor supply) are regarded as induced effects as firms are generally not the primary target group of the intervention.

Figure 1:
graphical illustration of employment-related effects of development interventions.



(1) Mediators: (e.g.) Productivity, Employability, Knowledge,... (2) Indirect Effects: (e.g.) Spillover effects, improved inputs, substitution, multiplier effects... (3) Induced Effects: (e.g.) Spillover effects, business environment effects, displacement (4) Feedback loops: Second- or Third-Order Effects (5) Externalities
Source: Authors.

and induced effects, depending on whether the subjects they affect are part of the target group or not. In general they are going to be a combination of both.

One could integrate feedback loops (sometimes called second- or third-order effects) within this conceptual framework: the changes in (employment) outcomes among beneficiaries or the target group that are induced by indirect or induced effects. For conceptual clarity, we refrain from a detailed discussion of these effects for now. Capturing also these (second/third-order effects) on the target group is often a matter of long-term data collection. However, the establishment of any causal link is methodologically very challenging. Finally, we also refrain from discussing the influence of (unintended) externalities, which occur if individuals outside the target group are affected by the outputs of an intervention.

Eventually, the overall employment *impact* on the economy depends on the entire set of employment effects – both inside and outside the target group of an intervention.[7] The second step of the methodological framework aims to quantify this overall employment impact of a program, taking into account also indirect and induced effects which cannot be estimated as part of a counterfactual analysis in the first stage.[8] This is important, since indirect and induced effects may be much more relevant than direct effects for some intervention types, as discussed in Section 2.2.

Note that the definitions of different employment effects and impacts established in this paper aim to be universally applicable to the various types of intervention commonly found in (German) development cooperation (see next section). Consequently, it needs to be more flexible than any intervention-specific terminology can be, in order to embrace a broad set of heterogeneous development programs within a common conceptual framework.

Box 1: Terminology of the intervention logic regarding employment

Outputs: The products and services which result from the completion of activities within a development intervention. A typical output of a human capital development program is the provision of training to participants of the program.

Outcome: The short-term and medium-term *level* of an outcome variable of interest caused by an intervention's outputs (sometimes called "result"). Employment-related outcomes are (e.g.) the employment situation of an individual or the number of employees in a firm.

(Treatment) Effects: All intended or unintended *changes* in outcome variables (directly or indirectly) brought about by the outputs of an intervention. Among these are employment effects as well as other (non-employment) effects. Depending on the population in which these changes take effect, one can distinguish between:

Direct employment effects: Changes in employment outcomes *among beneficiaries* of the intervention which are directly caused by the outputs of an intervention. Typically, direct effects are a primary goal of the intervention.

Indirect employment effects: Changes in employment outcomes *among the target population* which are caused by direct (employment and non-employment) effects of the intervention. Indirect effects comprise (e.g.) multiplier effects, substitution effects, or effects of altered policy and regulatory framework conditions.

Induced employment effects: Changes in employment outcomes among individuals and firms which are not part of *(i.e. outside) the target group* of an intervention. The effects are induced through the entire initial economic effect of an intervention – including direct and indirect effects. These include employment effects along the value chain (multiplier effects), effects of an altered business environment, or displacement (see Box 2).

7 This is in contrast the terminology sometimes found in evaluation studies, where the (causal/net) effect on the beneficiaries is already called the impact of an intervention (which we refer to as the net effect).

8 Whether an impact evaluation captures also indirect/induced effects depends on the design and data collection. This should be discussed in Section 3.1.

Employment impacts: The overall (economy-wide) change in employment as a result of the entire set of employment-related effects brought about by an intervention. Employment impacts reflect the additional employment creation that can be attributed to an intervention. Hence employment impacts are by definition always measured net of the potential counterfactual scenario or any negative indirect/induced effects of interventions.

2.2 Preparatory step (i): classification of intervention types

Development projects vary considerably in their potential to affect the employment situation of the target population. This does not only concern the number of jobs potentially created through an intervention but also the *type* (e.g. self/dependent, formal/informal, permanent/temporary, etc.) and *attributes* of the jobs created (e.g. hours, skill-level, payment, provision of benefits, target population, etc.). Moreover, the characteristics of the effect itself can differ across interventions: employment effects can differ regarding the type (e.g. direct/indirect) and timing of effects (e.g. short-/medium-/long-term).

Despite such heterogeneity, a detailed description of the results chain for each intervention regarding employment effects may not be feasible or informative as part of a portfolio analysis. A systematic framework for assessing employment effects for a portfolio of development projects therefore requires some form of a-priori classification of development interventions that adequately reflects this heterogeneity in the characteristics and size of potential employment effects.

Such a classification will facilitate an assessment of the job-creation potential for a larger portfolio without detailed knowledge of each project's specific intervention logic. Furthermore, the classification can also provide the basis for a detailed portfolio analysis because the particular characteristics of employment effects have important implications when choosing which projects of a portfolio to evaluate – and how to design the evaluation (see next section). Hence, the first step in a portfolio assessment should provide a systematized overview of expected employment effects for each program of interest.

The analysis of employment-specific results chains of German development interventions (see RWI 2013) shows, however, that the usual practice – followed by many development agencies – to structure the portfolio along thematic sectors bears only little informative value about potential employment effects (clearly this is because typically other issues take center stage). Irrespective of the sector, interventions often bring about widely different employment effects in terms of size and characteristics (type and timing). Moreover, development agencies increasingly follow a multi-level approach to address complex issues within specific sectors. A classification of entire development programs in sectors will not be able to account for the heterogeneity across types of interventions, which is generally larger than across sectors.

We therefore propose a typology of development interventions that is based on the intervention-specific approach leading to employment effects. In particular, the classification reflects the particular barrier to employment generation that is addressed through the outputs of development interventions. Accordingly, we distinguish between five different intervention types: (i) Human Capital Development; (ii) Private Sector Development; (iii) Infrastructure Development; (iv) Institutional Development; (v) Policy Advisory.

Table 1 presents typical outputs of these interventions and characterizes employment effects that can be expected at various levels (direct, indirect, and induced). The table is meant to provide an overview but not an exhaustive list of all potentially employment-related effects that a detailed results chain for a specific project could deliver. A more detailed discussion and some guidance on how individual interventions may be systematized along this classification can be found in RWI (2013).[9] As a point of reference, the

9 This table is an extended table, based on that found in the original study. Hence, while the classification of intervention types remains the same, the description of employment effects has been developed further according to the refinement of the terminology presented in Section 2.1. Most importantly, the typical employment effects in Table 1 are formulated to reflect the quantitative and qualitative dimension of employment (see Section 5.1).

type of intervention can often be inferred from the type of service delivery provided by the intervention and the beneficiaries who are typically affected by the output of the intervention.[10] As an example one may cite different approaches of financial sector development: Private sector development projects would support existing banks by (directly) providing business services. Institutional development projects would train individuals in the institutions to provide better services to the banking sector. Policy advisory projects would aim to support politicians on possible reforms of financial sector regulations.

Box 2: Terminology of employment-relevant indirect and induced effects

Several types of indirect and induced effects are of particular relevance in the context of employment analyses. As no agreed-upon definitions of these exist in the literature, we propose to use the following concepts according to the classification of employment effects above:

Substitution effects: Employment effects within the target population at the expense of other subjects within the target population that did not benefit from (participate in) the intervention.

Displacement effects: Positive employment effects within the target group are offset by negative employment effects for individuals/firms outside the target population.

Multiplier effects: Employment effects as a consequence of the initial economic effect of the intervention (direct or indirect). This includes supplier-effects along the value chain as well as effects of an increase in consumption (consumption multiplier).

The above effects can generally be considered as specific forms of general equilibrium effects (factor prices, demand). Additionally, direct effects of interventions may have employment-relevant effects on the conditions under which individuals and firms operate. We subsume these as

Business environment effects: Encompass a variety of effects, which generally affect the productivity or transaction costs of firm and/or individuals on the market. This includes (i) improved policy or regulatory frameworks in product- and factor markets (incl. labor markets); (ii) new and improved market inputs, including improved supply of qualified labor; (iii) new and improved public goods (infrastructure) and public service delivery (e.g. education).

Since these intervention types also reflect general approaches to addressing development challenges, this typology could easily be employed to systematize a portfolio in a broader (not employment-specific) development context. However, the primary objective of this systemization is to establish a framework for analyzing which types of interventions are most effective in generating employment in which contexts and sectors, based on impact evaluations using a comparable indicator.

10 As mentioned in the previous section, the subjects addressed by the output of an intervention do not necessarily equal the target group or final beneficiaries of an intervention: For example, an intervention may work with employees of labor-market institutions in a region but ultimately target the rural poor by improving job-placement services offered to them.

Table 1. Classification of intervention types from a job-creation perspective

	Typical Outputs	Direct Employment Effects (beneficiaries)	Indirect Employment Effects (within target group)	Induced Employment Effects (outside target group)
Type (i): Human Capital Development	Education, Training and Health Interventions	(+/-) individual-level via *employability, entrepreneurship skills (e.g.)*	(+) individual-level via *knowledge spillovers* (-) individual-level via *substitution effects*	(+) individual-level via *knowledge spillovers* (-) individual-level via *displacement* (+) firm-level via *business environment effects (increased supply of qualified labor , productivity)* (+) firm-level via *multiplier effects (consumption)*
Type (ii): Private Sector Development	Business services, credit supply (amongst others)	(+/-) firm-level via *productivity, new firms (e.g.)*	(+) firm-level via *business environment effects (input supply, increased consumption)* (+) firm-level via *value chain / multiplier effects (supplier effects, consumption effects)* (+) firm-level via *knowledge spillovers* (-) firm-level via *substitution effects*	(+) firm-level via *business environment effects (input supply, increased consumption)* (+) firm-level via *value chain / multiplier effects (supplier effects, consumption effects)* (+) firm-level via *knowledge spillovers* (-) firm-level via *displacement*
Type (iii): Infrastructure Development	Monetary and non-monetary resources, consulting	(+) temporary, firm-level via *construction and operation* (+/-) firm-level via *productivity (e.g. input supply)* (+/-) individual-level via *productivity (e.g. public goods)*	(+) firm-level or individual-level via *business environment effects (framework conditions (infrastructure), input supply)* (+) firm-level via *value chain / multiplier effects (supplier effects, consumption effects)* (+) firm-level and/or individual-level via *knowledge spillovers* (-) firm-level or individual-level via *substitution effects*	(+) firm-level and/or individual-level via *business environment effects (framework conditions, input supply)* (+) firm-level via *value chain / multiplier effects (supplier effects, consumption effects)* (+) firm-level and/or individual-level via *knowledge spillovers* (-) firm-level and /or individual-level via *displacement effects*

Table 1. (continued)

	Typical Outputs	Direct Employment Effects (beneficiaries)	Indirect Employment Effects (within target group)	Induced Employment Effects (outside target group)
Type (iv): Institutional Development	Organi-zational development, monetary and non-monetary resources, training	(+/-) firm-level via *productivity (e.g. regulatory framework)* (+/-) individual-level via *employability (e.g. public service provision)*	(+/-) firm-level or individual-level via *business environment effects (framework conditions, regulatory and institutional mechanisms)* (+/-) individual-level via *framework conditions (public service delivery (education), regulatory and institutional mechanisms)* (+) firm-level via *value chain / multiplier effects (supplier effects, consumption effects)* (+) firm-level or individual-level via *knowledge spillovers* (-) firm-level or individual-level via *substitution effects*	(+/-) firm-level and/or individual-level via *business environment effects (framework conditions, input supply, qualified labor)* (+/-) individual-level via *framework conditions (public service delivery, regulatory and institutional mechanisms)* (+) firm-level via *value chain / multiplier effects (supplier effects, consumption effects)* (+) firm-level and/or individual-level via *knowledge spillovers* (-) firm-level or individual-level via *displacement effects*
Type (v): Policy Advisory and Awareness Raising	Advisory, Research, Workshops, Political Campaigning	(+/-) firm-level via *productivity, new firms (e.g. policy framework)* (+/-) individual-level via *employability (e.g. public service provision)*	(+/-) firm-level and/or individual-level via *business environment effects (framework conditions, policy environment, regulatory mechanisms)* (+/-) individual-level via *framework conditions (public service delivery (education), regulatory and institutional mechanisms)* (+) firm-level via *value chain / multiplier effects (supplier effects, consumption effects)* (+) firm-level or individual-level via *knowledge spillovers* (-) firm-level or individual-level via *substitution effects*	(+/-) firm-level and/or individual-level via *business environment effects (framework conditions, input supply, qualified labor)* (+/-) individual-level via *framework conditions (public service delivery, regulatory and institutional mechanisms)* (+) firm-level via *value chain / multiplier effects (supplier effects, consumption effects)* (+) firm-level and/or individual-level via *knowledge spillovers* (-) firm-level or individual-level via *displacement effects*

Source: Authors, based on RWI (2013).

Table 2. Classification to assess the evaluability of interventions ex ante

Expected employment effect	Direct			Indirect / Induced		
Target group are...	Beneficiaries	Individually identifiable	Difficult to identify	Beneficiaries	Individually identifiable	Difficult to identify
Human Capital Development						
Private Sector Development						
Infrastructure Development						
Institutional Development						
Policy Advisory and Awareness Raising						

Colors indicate the efforts typically needed in terms of data collection and methodological designs. Green: Low, Yellow: Medium, Red: High

Source: Authors, based on RWI (2013).

2.3 Preparatory step (ii): systematic project selection

Most development agencies face limitations regarding the resources they can devote to monitoring and evaluation of their programs. As a consequence, not every project may be (rigorously) evaluated as part of a portfolio analysis, especially in large portfolios. In some cases, the set of interventions that should be part of the portfolio analysis may already be determined by their size or political relevance a-priori. If this is not the case, evaluators should make an informed and transparent choice of the interventions which will be assessed in more detail as part of the portfolio analysis.

We propose a systematic approach for identifying those projects for which a more detailed assessment of employment effects is likely most informative and feasible. The approach is based on two features of development interventions: the first aspect to consider regards the potential magnitude of employment effects – which we refer to as "relevance". The second aspect is whether these employment effects can be evaluated: their "evaluability" in terms of efforts and resources needed and the robustness of results that can be expected.

The first aspect, the employment-relevance, requires an ex-ante appraisal of the potential magnitude of employment effects for each intervention in the portfolio. While a precise judgment is generally difficult[11], it may be best inferred from the type of intervention, as discussed in the previous section. The second aspect, the evaluability of employment effects, depends on whether the subjects (individuals or firms) that benefit from the intervention, the "beneficiaries", are easily identifiable among the target group of that intervention.[12]. In addition, an intervention's evaluability is likewise connected to the type of employment effect (direct/indirect/induced): Typically, the further down along the causal chain employment outcomes are effectuated, the more difficult is their measurement and attribution to the intervention. Interventions that generate direct, short-term employment effects are typically easier to evaluate.

As it turns out, there is a certain connection between the type and the evaluability of interventions. Table 2 depicts a possible approach to implement this idea of connecting the relevance and evaluability of employment-related develop-

11 In Section 5, we describe a process through which these ex-ante estimates could be continuously improved within a system of institutional learning about employment effects.

12 The beneficiaries can be considered as a subset of the target group of an intervention that is actually affected by the outputs of an intervention. For example, the target group of a training program may be young females while the beneficiaries are only those females who partake in the program.

ment interventions to inform an (ex-ante) selection of interventions.
In short, the nexus between the type and evaluability of interventions can be exemplified by two contrasting cases: For interventions with direct contact to the beneficiaries, such as most training programs (top left), it is generally feasible to collect individual-level data and to implement appropriate evaluation designs. To evaluate interventions for which even the target group is difficult to define, such as some national-level political advisory interventions (bottom right), even finding adequate secondary data may be demanding. Similarly, interventions that influence employment outcomes indirectly and in the long run will be more difficult to evaluate. Hence, being able to judge the required effort and benefits of a rigorous evaluation is advisable.

As an example to demonstrate why the proposed classification may be more adequate for project selection than solely the sector of projects, compare a political advisory project on environmental issues with a training program for unemployed youth: the former, for example inducing a change in the regulatory framework for garbage collection, may generate or improve a greater number of jobs than the youth training program – but these effects likely differ systematically in their characteristics, with important implications for the evaluability of interventions.

Of course, other aspects may influence the final choice of projects as well, e.g. their size or geographic location. But, while it may be interesting to evaluate projects with only minor employment effects to know about their potential for upscaling or cost-benefit comparisons, the estimates will not affect the aggregate figure (and its precision) much. The criteria proposed above aim to serve as a guideline to select a portfolio such that the aggregate figure comprises the most precise estimates for the most employment-relevant projects. Hence, this approach should be considered in the context of assessing aggregate effects for a portfolio, and implies a focus on conducting evaluations that are least costly and easiest to implement in order to arrive at the best estimate for the aggregate impact. In contrast: Section 5 (Step 3) argues that it may be of particular relevance in some instances to deliberately select interventions for which an evaluation is challenging, in order to generate information to feed into the learning system.

Finally, the proposed classification – based on the intervention type, the possibility to identify beneficiaries, and the type of employment effect (direct/indirect/induced) – can be carried out even in the design phase of programs (ex-ante). This would allow an early integration of adequate M&E Systems in the design of employment-relevant programs. Generally, the earlier an impact evaluation of employment outcomes is taken into account, the less effort is required to arrive at rigorous impact estimates.

Typically, indirect effects are more difficult to evaluate than direct effects since they cannot be measured among the beneficiaries. The beneficiaries of an intervention are often clearly identified as those receiving the outputs of an intervention (and hence are often “within reach” of the program). Its target group may be more or less precisely defined, depending on the intervention. In order to measure changes in outcomes among the target group, data collection has to clearly define and identify the individuals affected by the intervention. This often depends on whether the intervention aims at a specific target group that is clearly distinguishable (identifiable) among the larger population. At the same time, indirect effects may be of high relevance for the evaluation, if (positive or negative) effects on the target group are of importance to judge the success of the development intervention.

Induced effects gradually unfold through the entire economy and often do not pertain to a specific set of firms or individuals. Consequently, data collection will generally require larger and more complicated sampling to capture all of the possible effects. Hence, indirect employment effects of an intervention will still be easier to evaluate than induced effects as data collection generally requires less effort and the effects are more closely linked to the intervention, with less assumptions needed along the theory of change Therefore, a precise empirical estimation of induced effects does not only require intense data collection efforts but is also methodologically challenging.

3. Step 1: estimating net employment effects at the intervention level

The first step in the bottom-up procedure – after the portfolio has been determined for which the aggregate employment impact is to be assessed – is to estimate net employment effects for each intervention in the portfolio. This section describes two approaches to arrive at such an estimate, each varying in terms of efforts/resources needed and the quality/precision of results to be expected.

Employment effects of an intervention can be measured in gross or net terms: Gross employment effects simply compare the employment outcome of beneficiaries before and after the intervention (e.g. the number of employees), thus implicitly assuming that the employment situation of beneficiaries would have remained unchanged, had the intervention not gone underway. Arguably, this is an unrealistic assumption in many cases, if not in most cases. To determine the success of development projects in improving employment outcomes of beneficiaries, we are ultimately interested in the *net* effect of the intervention: The realized employment outcome net of the employment outcome that would have occurred even in the absence of the development project (i.e. the number of *additional* jobs that have been created). To arrive at net employment estimates thus requires the estimation of the *counterfactual* scenario. This methodological challenge is presented in more detail in the next subsection.[13]

3.1 First-best option: counterfactual impact evaluation

The objective of an impact evaluation is to estimate quantitatively the causal effect of the intervention (treatment) on the outcome the program wants to influence. Modern evaluation research has come to utilize a counterfactual concept of causality, which in several steps of methodological development over the last decades has taken on the shape in which it is used today (Holland 1986). This model defines the causal effect of a treatment as the difference between the factual outcome ("Of the 100 training participants x per cent found a job") and the counterfactual case ("What percentage of the same 100 training participants would have found a job without the program?").[14]

Clearly, the counterfactual is a hypothetical construct and can never be observed in data, since no individual or group can be both exposed to the intervention and not exposed to the intervention at the same time. Holland (1986) refers to this as the fundamental problem of causal inference. In order to evaluate the effect of the treatment, we therefore always need to compare distinct units receiving the different levels of the treatment. Such a comparison can involve different physical units or the same physical unit at different times (Imbens and Wooldridge 2009). The key conceptual point, however, is that both factual and counterfactual states are defined for the individual unit, along with realizations of the outcome variable for each state. For this reason, this causal model has become known as the Potential Outcome Model.

The basic notation is as follows:

- Treatment indicator: D (=1 if participating / exposed to the intervention; =0 if not participating / not exposed to the intervention). In general, D={0,1,...,T}, most applications focus on the binary case, D={0,1}, i.e. a treatment state and a no-treatment state.
- Outcome Y. (E.g. the employment situation of an individual or the number of employees at the firm-level.)
- Observable characteristics (socio-demographic characteristics) that may influence treatment assignment and the outcome are captured in the vector X. The covariates X are pre-determined, i.e. they are not influenced by participation.
- Individuals (units of observation): i = 1,...,N

The potential outcome model in the binary case (treatment: yes/no) is therefore:

$Y = Y_0 \quad if \ D = 0,$

$Y = Y_1 \quad if \ D = 1,$

13 Note that this methodological challenge is concerned with the correct *measurement* of employment impacts to which the intervention has contributed. It does not specifically address the question of attribution.

14 We are explicitly distinguishing between the "counterfactual" and "deadweight". The latter usually includes the likelihood that a project or policy change would have also been realized (e.g. funded), had it not been initiated by the development agency. This is of little interest for the study of employment effects, and rarely possible to quantify. We restrict our interest to the counterfactual scenario that replicates the result of a counterfactual evaluation.

This illustrates the missing data problem that impact evaluation research faces (in the binary case):

	D=1 (participants)	D=0 (non-participants)
Y_1	Observable	Unobservable (counterfactual)
Y_0	unobservable (counterfactual)	Observable

The model has several important implications. First, the causal effect at the individual level cannot be observed. Hence, we need an adequate parameter to summarize individual causal effects, and empirically impact evaluations will focus on the estimation of average causal effects. Various types of averaging are possible in this regard; the most commonly used evaluation parameter is the Average Treatment Effect on the Treated (ATET):

$$\text{ATET: } E(\Delta \mid D = 1) = E(Y_1 - Y_0 \mid D = 1) = E(Y_1 \mid D = 1) - E(Y_0 \mid D = 1)$$

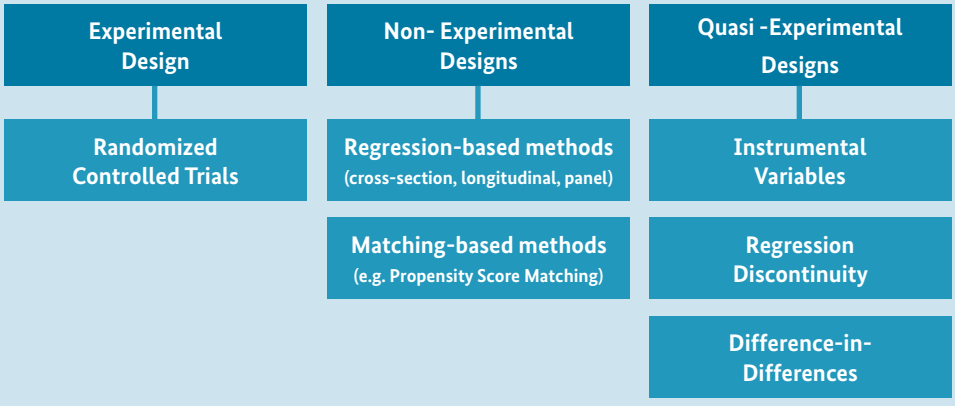
where E(.) denotes the expectations operator, and the second term in the last equation is the counterfactual (= the average outcome of the participants in the case of non-participation). This counterfactual $E(Y_0 \mid D = 1)$ is, in econometric terms, not identified. A parameter is identified, if it can be estimated with ever increasing precision as the sample size increases. This, however, is the case for the first term $E(Y_1 \mid D = 1)$, which is identified from observable data and could be estimated with perfect precision if N were infinitely large. The evaluation problem is therefore: What assumption allows replacing the unobservable counterfactual average by an alternative, observable population average? Such an identifying assumption, if correct, allows for construction of a counterfactual situation and identifies the population parameter. Note that due to the unobservable nature of the counterfactual every impact evaluation requires one or more identifying assumptions. These must be justified and made plausible by the researcher, since they cannot be statistically tested (at least not fully), hence cannot be right or wrong a priori nor proven right or wrong a posteriori.

This – simple and well-known – delineation of the causal model underlying impact evaluations is explained here, because it facilitates the description of the various methods that can be used to rigorously estimate employment effects of interventions. These methods are competently and comprehensively reviewed in several articles and books, such that a detailed explanation is beyond the scope of this paper. Important resources in this regard include, for instance, articles focusing predominantly on the econometrics of impact evaluation (e.g. Heckman et al. 1999, Imbens and Wooldridge 2009) and several books providing guidance for practitioners (e.g. Gertler et al. 2011, Khandker et al. 2010).[15] It is worth noting that many of the empirical methods for causal analysis now commonly used in evaluation research have been developed explicitly in the case of evaluating training and other employment programs (Imbens and Wooldridge 2009).

15 A more concise guideline to measuring employment effects with explicit connection to GIZ interventions is Kluve (2011).

Figure 2.

Classification of impact evaluation designs



The key virtue of formulating the problem of causal inference using the Potential Outcome Model is that it lays open the central role of the relationship between treatment assignment and potential outcomes. Moreover, this allows to correspondingly to classify the set of available methods into three groups, following Imbens and Wooldridge (2009). Methods for causal inference based on (1) randomized experiments; (2) assuming selection on observable characteristics, and (3) selection on unobservables. We will consider these three categories and the methods they comprise in turn. Figure 2 gives an overview of the main methodological tools available for program evaluation.

3.1.1 Randomized experiments

The most straightforward case for analysis is when assignment to treatment is randomized (in a controlled way by an experimenter) and, thus, independent of covariates X as well as the potential outcomes Y. In such classical randomized controlled trials (RCTs) it is relatively easy to obtain estimators for the average effect of the treatment, using e.g. the simple difference-in-means by treatment status. Randomized experiments have been used in the evaluation of labor market programs since the 1970s (starting in the US), with some increasing trend over the last decade, though still not at a very large scale. In recent years, RCTs have increasingly been used in development economics.

3.1.2 Selection on observables

More common is the case in which researchers analyze data from non-experimental (also called: observational) studies. Non-experimental data generally create challenges in estimating the causal effect of programs. But in one important special case, questions regarding identification and estimation of the policy effects are rather well understood (Imbens and Wooldridge 2009). This case is variously referred to as selection-on-observables, unconfoundedness, exogeneity, or ignorability. All these labels refer to some variant of the same assumption: That adjusting the treatment and control group for differences in characteristics observed prior to treatment (covariates X) removes all biases in a comparison of outcomes Y between treated and control units.

This case is of great practical relevance, with many impact evaluation studies relying on some form of this assumption: Specifically, this category comprises classical regression methods, e.g. adjusting for covariates X in a linear regression. Another method that is based on the unconfoundedness assumption and has been applied increasingly often is "Statistical Matching", generating samples of treated and comparison units that are balanced in their observed X. This approach is thus trying to mimic a randomized experiment ex-post. In practice, in recent years the most frequently used version of a selection-on-observables design has been Propensity Score Matching, adjusting for the (estimated) conditional probability of receiving the treatment given the observed characteristics (covariates X).

3.1.3 Selection on unobservables

Without unconfoundedness, there is no general approach to estimating treatment effects but various methods have been proposed for special cases (cf. Imbens and Wooldridge 2009). These approaches are often associated with the concept of "natural experiments", in which policy changes (or other "exogenous shocks") can be used to effectively define (randomly assigned, though not in a controlled way) treatment and control groups. Three of them are of major importance for empirical practice. One method is the instrumental variables approach that relies on the presence of additional "treatments", the so-called instrumental variables. Essentially, in the case in which treatment assignment is endogenous (i.e. systematically connected with the potential outcomes), researchers look for instrumental variables that satisfy two assumptions: First, the instrument is correlated with the treatment (testable assumption), and second, the instrument does not exert a direct impact on observed outcomes, but only through the treatment (maintained, untestable hypothesis). A second method is the regression discontinuity design (RD) that applies to settings in which (in its pure form, the so-called "sharp" RD) overlap is completely absent because the assignment is a deterministic function of one or more covariates, but causal comparisons can be made exploiting continuity of average outcomes as a function of the covariates. (In the "fuzzy" RD design the assignment probability does not switch from 0 to 1 as in the sharp design, but only requires a (sufficiently large) discontinuity in the probability of

treatment assignment at the threshold determined by the forcing covariate(s)).

Finally, the third method, difference-in-differences (DiD), relies on the presence of additional data in the form of samples of treated and control units before and after the treatment (these can be panel data or repeated cross-sections). In the simplest setting outcomes are observed for units in one of two groups, in one of two time periods. Then the average gain over time in the control group is subtracted from the gain over time in the treatment group. This double differencing removes biases in second-period comparisons between the treatment and control group resulting from permanent differences between the groups, as well as biases from comparisons over time in the treatment group resulting from time trends unrelated to the treatment. The intuitive way in which the DiD design can remove important biases along with its broad applicability in many different contexts has made this method one of the most frequently applied designs to estimate causal effects. Nonetheless, in practical applications attention must be paid to challenges to the design (e.g. sensitivity of estimates to the timing of measuring outcomes; time trends differentially affecting treatment and control groups, etc.).

Regarding data availability and collection, the experience shows that existing monitoring systems of development programs often do not provide the information required to rigorously estimate employment effects using the above methods of counterfactual impact evaluation.[16] Furthermore, in many developing countries, secondary data sources are often not sufficiently detailed or readily available to evaluate programs in the short-term. Hence, when called to analyze employment effects ex-post (i.e. when the program is already underway or finished), evaluators are sometimes not able to implement rigorous evaluation designs.

16 See the discussions in (e.g.) Riddell (2014) or Norad (2014), as well as the experiences documented in the pilot study on Morocco underlying this paper (RWI 2013).

3.2 Second-Best Option: Approximation using Gross Employment Effects

The goal of the original research project (RWI 2013) was to delineate approaches that allow providing an estimate of the overall employment impact, i.e. aggregated over all projects in a given portfolio of development interventions within one country. As discussed above, for various reasons, it may not be possible to implement a rigorous design for every project as a basis for aggregation within the bottom-up procedure. Some of these projects, however, may be of particular relevance for the portfolio. In order to provide the possibility to integrate these projects into the aggregate impact estimate, we discuss a second-best option to approximate net employment effects when rigorous designs are not feasible.

The main (and simple) idea of this second-best procedure is to first estimate *gross* employment effects of interventions within existing monitoring systems (e.g. via simple before/after comparisons). Then, in a second step, a parameter for the assumed counterfactual of the specific intervention is quantified and applied to convert gross employment effects into net effects, approximating the latter. Information specifying this counterfactual parameter can come from previous (external) studies which evaluated similar programs in similar contexts.

Consider an impact evaluation study from which we want to derive a value for the counterfactual parameter. The counterfactual is represented by the related estimate for $E(Y_0 \mid D = 1)$ and hence the counterfactual parameter P^C can be derived as

$$P^C = \frac{E(Y_0 \mid D = 1)}{E(Y_1 \mid D = 1)} = \frac{E(Y_1 \mid D = 1) - ATET}{E(Y_1 \mid D = 1)};$$

where ATET represents the (net) treatment effect as reported in the evaluation study and $E(Y_1 \mid D = 1)$ is the mean outcome in the treatment group. The parameter thus represents the share of the observed outcome in the treatment group that would have been realized even under the (hypothetical) counterfactual scenario. Hence, lower values of the parameter represent a higher net impact.

Say, for example, an evaluation conducted an RCT and finds the average increase in the number of employees in participating firms after an intervention is $Y_1 = 10$ and the average among the control group is $Y_0 = 3$. The implied counterfactual parameter would be $P^C = (10 - (10 - 3))/10 = 0.3$. One can apply this parameter to the results of follow-up intervention which measured an average gross-effect of 15 employees (e.g. via a before-after among participating firms). The net employment effect per firm can be computed as *Net-Effect=Gross-Effect ×(1 −Counterfactual)= 15×(1 −0.3)= 10.5* (cf. section 5.1). If 500 firms participated, the implied net employment effect would be 5250 net additional jobs.

Furthermore, within the learning system discussed in section 5, information from evaluations of previous phases or similar programs can be used to improve this estimate based on the knowledge continuously generated *within* the (German) development cooperation agency. Practically, every thoughtfully conducted impact evaluation delivers an implicit estimate of the counterfactual parameter that could be used to refine the initial approximations of counterfactual parameters. However, the results from studies should not be applied without an adequate assessment of the transferability, in particular an understanding of the comparability of the two interventions and their context. If no impact evaluation of a comparable intervention is available, it may be feasible to adjust the best-possible reference parameter based on qualitative information from program and country experts.

In addition to a credible value for the counterfactual parameter, the second-best approach requires a precise estimate of gross employment effects: That is, at least, the number of individuals or firms that have benefited from the intervention (number of beneficiaries) and the employment outcome of interest after the intervention, irrespective of their counterfactual (potential) outcome in the case they would not have been part of the intervention. In many situations it will be necessary to quantify a baseline (pre-intervention) and follow-up (post-intervention) measure of the employment-related outcome of interest.

Hence, to implement the second-best option still demands designing an adequate monitoring system to capture employment outcomes of beneficiaries. For certain intervention types even this may be challenging within typical M&E Systems, especially if employment effects are largely indirect or induced. Since baseline values are required in most cases, the monitoring design should be already included in the design phase of projects for which employment effects are of interest.

The proposal to quantify second-best estimates based on a counterfactual parameter is an integral part of the practical approach of this research project.[17] Nonetheless, it should be emphasized again that this procedure can only serve as an alternative "of last resort" to the results from a properly designed impact evaluation. The idea of the second-best is to provide an evidence-based and transparent procedure to arrive at an informed and plausible estimate, only in situations where an estimation of employment effects is otherwise not attainable and an aggregation including the specific intervention desired.

17 We also propose this second-best option in view of a critique often brought forward against academics advocating for rigorous impact evaluation designs (e.g. RCTs) when studying development effectiveness: That interventions which cannot easily be evaluated in a rigorous manner are consequently not evaluated at all; or – worse – that these type of interventions may not even get funding due to the pressure of proving results.

Back to Contents

4. Step 2: estimating economy-wide employment impacts

Section 2 above has already indicated that the initial effect of an intervention (estimated in the first step) will likely induce further economic effects. Many development programs only measure and report on their direct employment effects – yet their impacts in terms of indirect or induced effects on individuals outside the immediate target population may be more substantial. As outlined in Boxes 1 and 2, the initial effect of an intervention may trigger a large variety of effects on individuals and firms outside the primary group of beneficiaries. In combination, these indirect and induced effects can reinforce or reduce the immediate employment effect of an intervention: Substantive additional local consumption from increased earnings among beneficiaries may cause further job-creation. On the other hand, if workers in supported enterprises displace those outside the group of beneficiaries, it weakens the desired impact on aggregate welfare. These effects may be intended by the program or not – but their magnitude can certainly be influenced by the intervention design.[18]

For this reason, it can be argued that the overall, economy-wide employment impact (which comprises direct, indirect and induced effects) constitutes the relevant estimate of a meaningful evaluation regarding employment:

First, it is the policy-relevant estimate, since most development cooperation programs aim to improve the overall labor market situation (and hence welfare) in the partner country. Reporting only direct employment effects would miss the benefit of additional employment generated along the value chain (multiplier effects); or overstate the benefit of a program if it improves the employment situation for one group at the expense of another (displacement).[19] Hence, neglecting indirect and induced effects may lead to an over- or underestimation of the true employment impact.

Second, measuring also indirect and induced effects provides the best estimate to inform about the potential consequences of up-scaling or replication of programs, since it includes information about effects on individuals indirectly affected by the intervention – which likely gain in relevance the larger the initial effect and beneficiary group of an intervention.

Third, employment impacts deliver the only credible estimate by which programs can be meaningfully compared and aggregated across a broad range of intervention types. Most importantly, any conclusive Cost-Benefit-Analysis and comparison should be based on employment impacts: Measuring only direct effects may bias future portfolio designs towards interventions with large direct (but possibly short-lived) effects.[20]

Impact evaluation should therefore take into account possible ways of estimating these employment-relevant indirect and induced effects – however, the challenges regarding research design and data collection are substantial. The empirical designs and methods for data collection presented in Step 1 often only allow a partial analysis of employment effects. Only in specific circumstances this partial analysis captures indirect effects of interventions and generally does not comprise effects outside the target group. Even though experimental and quasi-experimental approaches to estimate indirect and induced effects have been developed, these designs are often data demanding and difficult to implement within commonplace evaluation designs.[21]

To account for the fact that in many cases it will not be feasible to collect data needed to assess these effects, this section introduces a practical solution to account for the (potentially large) influence of indirect and induced effects. The approach – similar to the Second-Best Option outlined in Section 4 – is to derive parameters about the relevance of these effects for specific intervention types from previous program evaluations and the existing literature. These parameter values can be applied to convert net-employment

18 Unfortunately, there is little systematic knowledge so far on how certain program design features influence the magnitude of specific indirect effects. This paper aims contributing to a systemized framework for further research and evaluation needed in this area.

19 Which Crépon et al. (2013) refer to as playing "a (costly) *musical chairs* game".

20 This is connected to the issue raised above, that programs generating employment effects that are difficult to measure (e.g. indirect, long-term effects) may risk not getting financed.

21 See, for example, Crépon et al. (2013)

effects (measured in Step 1) into net-employment impacts for the entire economy under study.

Unfortunately, as mentioned above, the evidence base about indirect effects is thus far comparatively weak. Hence, as long as impact evaluations are not more often geared towards an estimation of indirect/induced effects, parameter values will have to rely decidedly on the knowledge of thematic and country experts as well as stakeholders of the intervention. In particular, since the magnitude of indirect and induced effects depend strongly on the intervention and context characteristics. In RWI (2013) we discuss in more detail how project activities, outputs and existing studies can be systematically utilized to inform plausible parameters for specific indirect and induced effects. An important feature is that every output generated over the course of a program may potentially deliver additional information about employment effects. In any case, the process and information that enter the parameter estimate should be clearly documented: This allows refining the estimate once further information is available and ensures transparency regarding the assumptions underlying the estimated employment impact.

A complementary approach may be to systematically review findings from existing (impact evaluation) studies within and outside development cooperation to provide reference values of indirect and induced effects for specific intervention types. Table 3 displays example parameter values that originate from an *initial and preliminary search* of the literature, conducted as part of the original research project.[22] The parameter values are classified along the intervention types presented in section 2 and can be used to calculate lower and upper bounds of economy-wide employment impacts.

Parameter values in this table are presented as ratios of the net effect (as computed for the counterfactual parameter in Section 3.2) Hence, they are entering as (1-X) factors in the case of displacement and substitution and as (1+X) factors in the case of multipliers. The formula for converting net- or gross -effects into (economy-wide) employment impacts based on these parameter values can be written as:

Employement Impact

= Net – Effect x (1± – indirect Effects ± induced Effects)

= Gross – Effect x (1 – Counterfactual)
x (1 – Substitution)
x (1 – Displacement)
x (1 + Multiplier)

Within the "learning system" presented in the next section, these parameter estimates can be continuously refined as part of the ongoing experience generated within of development cooperation.

Table 3. Example parameter values for the calculation of indirect and induced effects

	Displacement	Substitution	Supplier	Consumption
Human capital development	0.1 – 0.3	0.2 – 0.4	0.05 – 0.1	0.3 – 0.5
Private sector development	0.2 – 0.4	0.3 – 0.5	0.3 – 0.5	0.2 – 0.5
Infrastructure development	0.1 – 0.3	0.15 – 0.3	0.5 – 0.7	0.3 – 0.5
Institutional development	0.3 – 0.4	0.3 – 0.6	0.3 – 0.5	0.2 – 0.4
Policy Advisory / Lobbying	0.3 – 0.4	0.15 – 0.3	0.2 – 0.5	0.2 – 0.4

Source: Authors, based on RWI (2013)

22 This search included: Greenberg et al. (2011), English Partnerships (2008), Centre for Strategy & Evaluation Services (2006), European Union (2013), Maré (2005), Bondonio & Martini(2011), Criscuolo et al (2012), National Audit Office (2003), Mouqué (2012). Hence, the table is meant to illustrate the procedure within our framework, but does not claim to present ultimately valid parameter ranges.

Back to Contents

5. Step 3: Aggregation and integration into a system of institutional learning

The previous sections have outlined a methodological approach that proceeds stepwise. Preparatory steps include a classification scheme that categorizes development interventions from an employment perspective, and a classification scheme that maps intervention type and target group to an assessment of the evaluability of the intervention (presented in Section 2.2).

Based on these preparatory steps, a portfolio-focused analysis of employment impacts would proceed in three main steps: First, rigorous assessment of net employment effects of the single interventions. Second, estimating employment impacts based on the results from Step 1, taking into account second- and third order effects. Step 3 now aggregates over all programs of the portfolio and discusses how estimates can be used to inform future ex-ante and ex-post assessments within a system that "learns" about employment effects.

5.1 Aggregation: a comparable indicator of employment effects

The aggregation of intervention-level employment impacts is straightforward once they have been estimated for each program in the portfolio. However, this requires that estimated values from Step 1 and 2 are comparable. The gradual implementation of a standardized indicator of employment effects in future programs would also allow to compare the effectiveness of different intervention types across sectors and countries. To this end, the introduction of a coherent indicator in programs that have been designated as relevant to employment creation would be necessary.

But employment comes in many dimensions. One may classify the dimensions of jobs by their *type* (e.g. self/dependent, formal/informal, permanent/temporary, etc.) and their *attributes* (e.g. hours, skill-level, payment, provision of benefits, target population, etc.). Both dimensions reflect the qualitative aspect of employment impacts. Normalized measures that cover both the quantitative and qualitative aspects of employment impacts are sometimes proposed – such as "the number of formal, full-time equivalent jobs per annum" created. However, these indicators are often data-intensive and bear little flexibility concerning the various aims and effects regarding employment outcomes of different programs. An adequate indicator, which can be applied to different (heterogeneous) interventions, should therefore be flexible as well as comparable across programs. It also needs to encompass both the quantitative as well as the qualitative aspects of employment effects.

We propose a simple indicator that comprises two sub-indicators, reflecting these two dimensions (quantitative/qualitative). It allows the comparison and aggregation both within each dimension as well as an aggregation into one single indicator. Furthermore, the proposed indicator can be adapted to different program objectives (e.g. target groups) without losing its general meaning and comparability. In particular, the qualitative indicator is deliberately defined flexibly, such that programs can formulate aspects of employment quality (type/attributes) that suit their interventions' goals.

The indicator is specified as follows:

1. The number of individuals whose employment situation has improved due to the intervention.

This indicator comprises two sub-dimensions:

a) The number of individuals additionally coming into employment, among whom are X women and X youths. (Quantitative indicator)[23]
b) The number of employed individuals whose employment quality or income situation has improved. (Qualitative indicator)

The indicator is aggregable and comparable across heterogeneous interventions, it avoids double counting and comprises all relevant dimensions of employment that are typically affected by development interventions.

5.2 Implementing a system of institutional learning about employment impacts

The systematic integration of a standardized indicator into employment-relevant projects not only facilitates an ex-post comparison and aggregation of employment effect of projects as part of a portfolio analysis. It also enables the

23 In recognition that job creation may not be relevant for some programs, measurement of this indicator is optional.

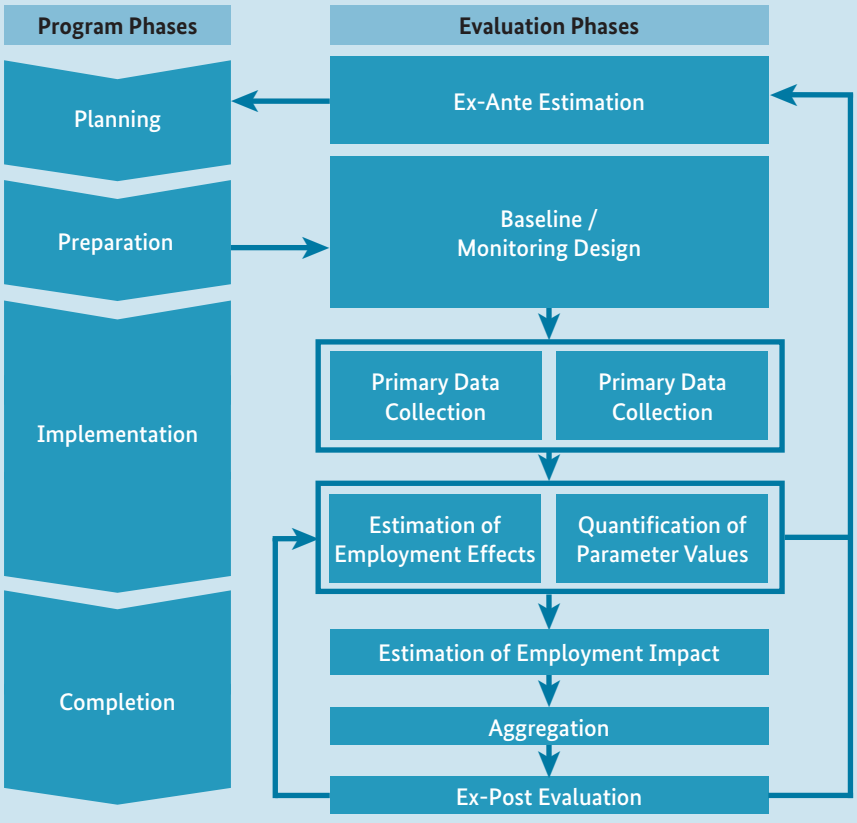
integration of a system of institutional learning about employment effects. This is the ultimate step in a systematically conducted portfolio analysis (geared towards sustainable knowledge creation).

As mentioned in the introduction, one objective of the original study (RWI 2013) was to provide a framework for an improved ex-ante appraisal of potential employment effects of development intervention – that is prior to program implementation. Plausible values can – and should – be derived based on previous experience and empirical evidence. Hence, the proposed framework specifically aims at an integration of project monitoring systems, ex-post evaluations and ex-ante assessments regarding employment impacts. That is, the system will allow any evaluation to inform ex-ante assessments of an upscaling or follow-up phase of the program. In practice, estimated parameter values from evaluations can be used to convert target values regarding (e.g.) the number of beneficiaries in follow-up phases into (more) plausible values for anticipated employment effects.[24]

Similarly, impact estimates of interventions can deliver a plausible parameter values that may be applied in an assessment of employment impact of comparable programs. Finally, the estimates generated in one application of the program can be used to validate or refine parameter values applied in previously conducted (second-best) evaluations. In a nutshell, every step conducted to estimate employment impacts of a portfolio within the learning system informs the subsequent step and validates the preceding step. This integrated process is depicted in Figure 3 below.

24 For instance, one could replace the number of participating firms that was measured ex-post in the example on page 24 by the target number of the follow-up phase to arrive at an ex-ante value for net employment effects.

Figure 3.
An integrated system for institutional learning about employment impacts



Back to Contents

If applied comprehensively within the work of development cooperation, this process systematically improves our knowledge of employment impacts of different intervention types via the continuous generation of comparable parameter values. Hence, the initial (rough) estimates of employment parameters (Table 3) can be continuously refined with each application of the approach. As part of this, empirically estimated parameter values could be capitalized to refine the upper and lower bounds displayed in Table 3.

The benefits of implementing a learning system are increasing with each application. That is, each evaluation that systematically feeds into the system lowers the resources and time required to determine credible/plausible employment impacts for similar programs or follow-up phases. In this regard, the implementation of a learning system will be more effective in delivering the desired benefits (in terms of delivering parameter values and subsequently reducing efforts), the more frequently M&E Systems measure employment effects in a comparable and coherent fashion – ideally via a standardized indicator as proposed in the previous section.

Obviously, another important aspect of this system is that high quality, rigorous impact evaluations are systematically conducted as part of the ongoing work of development cooperation. These impact evaluations should be designed to deliver comparable estimates of employment effects – including indirect and induced effects.

Back to Contents

6. Concluding remarks

In view of the rising prominence of jobs in the development discourse, donors look for solutions to the various challenges related to measuring employment effects of interventions. This paper aims to provide an input to the debate by outlining a framework through which employment impacts could be systematically assessed and, at the same time, generate the basis for a system of institutional learning about employment outcomes within development cooperation.

Benefits of the proposed bottom-up approach are manifold: The procedure provides an empirical estimate of the policy-relevant parameter – the economy-wide employment impact – based on a transparent and well-understood toolbox for measuring the effects of individual interventions. By generating comparable parameter values for indirect and induced effects, the system enables the inclusion of interventions which are typically difficult to evaluate into an (ex-post) portfolio assessment of development interventions. Moreover, the procedure allows to make use of the knowledge continuously generated within development cooperation in order to address data and resource limitations. In the long run, the efforts needed to provide a plausible estimate of aggregate employment impacts can thus be continuously reduced.

However, the learning system can only develop its potential benefits if adequate impact evaluations – that is: rigorous counterfactual designs based on a comparable indicator for employment outcomes – are systematically implemented in development cooperation programs in the future.

Back to Contents

References

Bondonio, D. and A. Martini (2012), Counterfactual Impact Evaluation of Cohesion Policy: Impact, cost-effectiveness and additionality, Final Report to DG Regional Policy "Counterfactual Impact Evaluation of Cohesion Policy", Associazione per lo Sviluppo della Valutazione e l'Analisi delle Politiche Pubbliche: Turin.

Centre for Strategy & Evaluation Services (2006), Study on Measuring Employment Effects – Final Report. Centre for Strategy & Evaluation Services (CSES): Sevenoaks.

Crépon; B., Duflo, E., Gurgand, M., Rathelot, R., Zamora, P. (2013), Do Labor Market Policies have Displacement Effects? Evidence from a Clustered Randomized Experiment. The Quarterly Journal of Economics, Oxford University Press, vol. 128(2), pages 531-580.

Criscuolo C., R. Martin, H. Overman and J. Van Reenen (2012), The Causal Effects of an Industrial Policy. NBER Working Papers 17842, National Bureau of Economic Research, Inc.

Donor Committee for Enterprise Development (2010), DCED Standard for Measuring Achievements in Private Sector Development – Control Points and Compliance Criteria, Version VI, January 2013. http://www.enterprise-development.org/page/measuring-and-reporting-results, last accessed Dec. 2nd 2014.

Fowler, B. and Markel, E. (2014), Measuring Job Creation in Private Sector Development. Working Paper for the Donor Committee for Enterprise Development (DCED). MarketShare Associates: Denver.

English Partnerships (2008), Additionality Guide: A standard approach to assessing the additional impact of interventions, third edition, English Partnerships: London.

European Union (2013), Expert evaluation network delivering policy analysis on the performance of Cohesion policy 2007-2013 – Job creation as an indicator of outcomes in ERDF programmes – Synthesis report, Report to the European Commission Directorate-General for Regional and Urban Policy, European Union: Brüssel.

Gertler, P. J., Martinez, S., Premand, P., Rawlings, L. B., and Vermeersch, C. M. (2011), Impact evaluation in practice. The World Bank: Washington, DC.

Heckman, J. J., LaLonde, R. J., and Smith, J. A. (1999), The economics and econometrics of active labour market programs, in Ashenfelter, O. and D. Card (eds.), Handbook of Labor Economics 3, Elsevier, Amsterdam.

Greenberg, D.H., G. Knight, S. Speckesser and D. Hevenstone (2011), Improving DWP Assessment of the Relative Costs and Benefits of Employment Programmes, U.K. Department for Work and Pensions Working Paper No. 100.

Holland, P. W. (1986), Statistics and causal inference (with discussion). Journal of the American Statistical Association, 81(396), 945-970.

International Labour Organization (2014). Global Employment Trends 2014: Risk of a Jobless Recovery? International Labour Organization (ILO): Geneva.

Imbens, G.W., and Wooldridge, J.M. (2009), Recent Developments in the Econometrics of Program Evaluation, Journal of Economic Literature Vol. 47(1), 5-86.

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

Kluve, J. (2011), Measuring employment effects of development cooperation interventions – Some methodological guidelines. Deutschen Gesellschaft für Internationale Zusammenarbeit (GIZ): Eschborn

Maré, D.C. (2005), Indirect Effects of Active Labour Market Policies, Motu Working Paper 05–01. Motu Economic and Public Policy Research Trust: Wellington.

Mouqué, D. (2012), What are counterfactual impact evaluations teaching us about enterprise and innovation support? Regional Focus Paper 2/2012, European Commission, Regional and Urban Policy: Brussels.

National Audit Office (2003), Regional Grants In England, The Department For Trade And Industry: Report By The Comptroller And Auditor General, Hc 702 Session 2002-2003, NAO: London.

Norad (2014), Can We Demonstrate the Difference that Norwegian Aid Makes? Norwegian Agency for Development Cooperation (Norad) http://www.enterprise-development.org/download.aspx?id=2424, last accessed Dec. 2nd 2014.

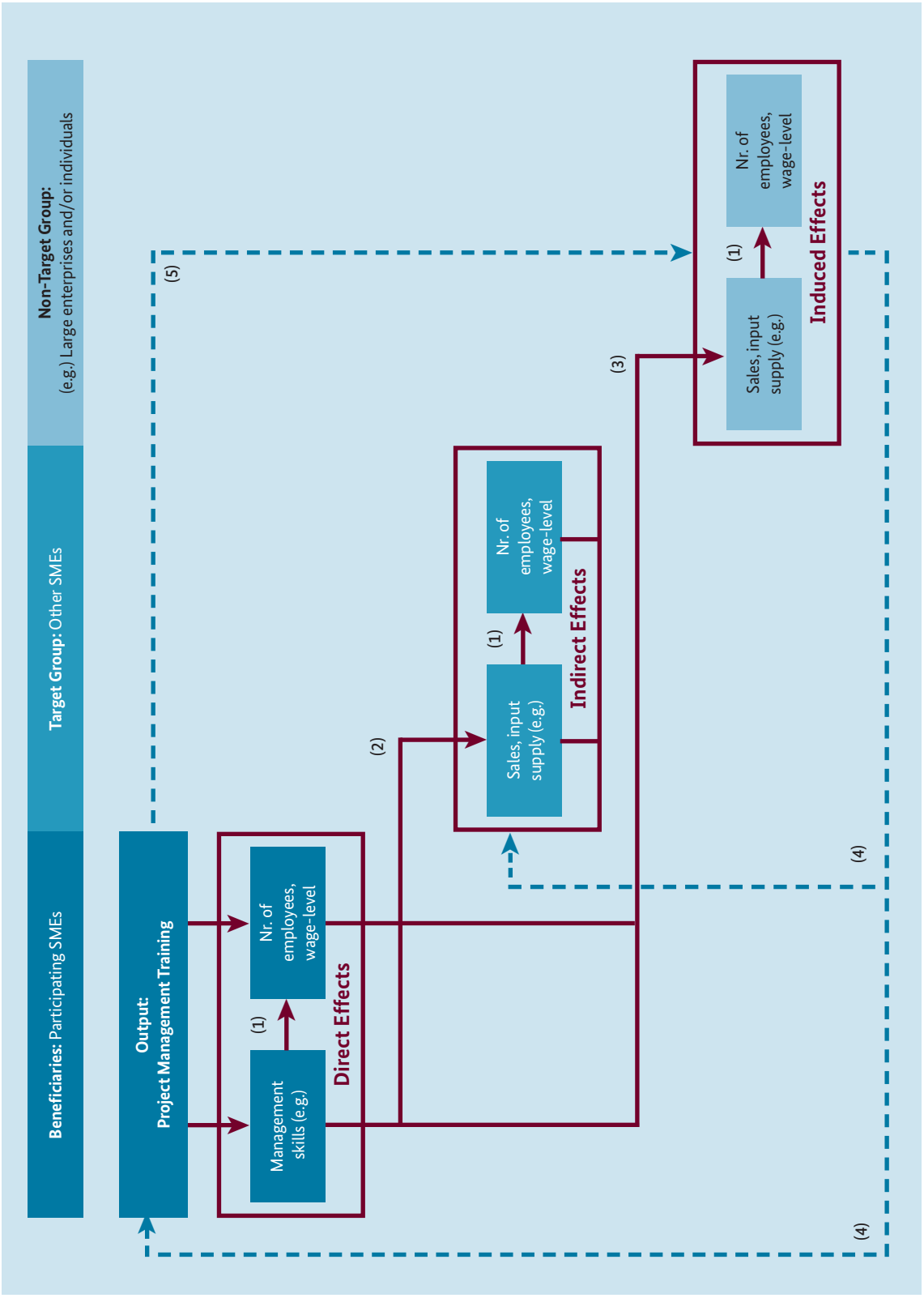
OECD-DAC (2002), Glossary of Key Terms in Evaluation and Results Based Management. Organisation for Economic Co-operation and Development (OECD): Paris.

Riddell, R. (2014), Does Foreign Aid Really Work? An Updated Assessment. Development Policy Centre Discussion Paper No. 33. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2409847, last accessed Dec. 2nd 2014.

RWI (2013), "Nachweis der Beschäftigungswirkungen von Maßnahmen der deutschen EZ – Pilotstudie Marokko" [Employment effects of German development cooperation interventions – pilot study Morocco], RWI Project Report: Berlin/Essen.

Appendix.

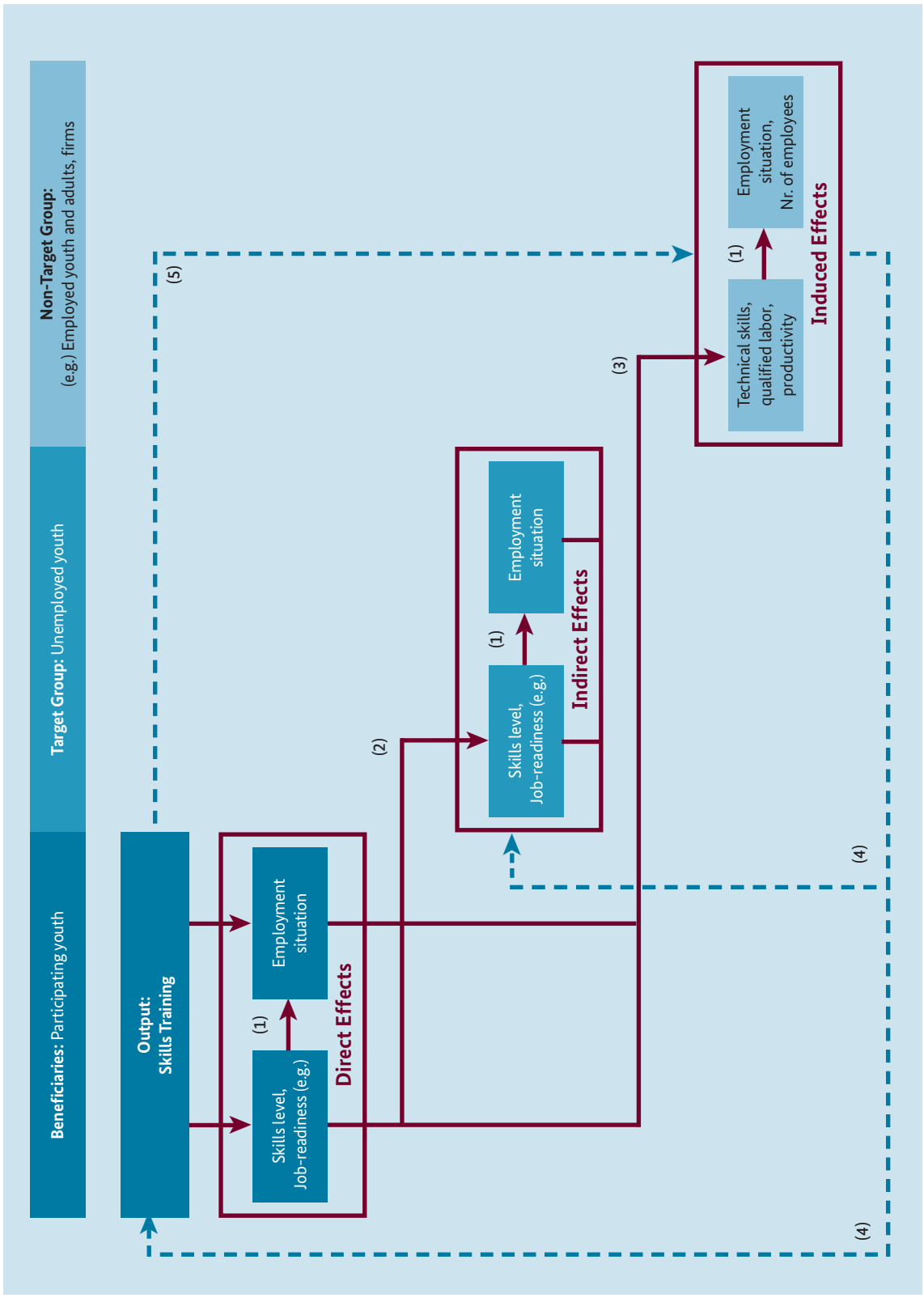
Figure A1: Graphical illustration of employment-related intervention effects. Example (i): SME promotion]



(1) Mediators: Productivity, Competitiveness, Sales... (2) Indirect Effects: Knowledge spillovers, substitution, multiplier effects... (3) Induced Effects: Knowledge spillovers, business environment effects (input supply), displacement,... (4) Feedback loops: Business environment effects, multiplier effects,... (5) Externalities.

Source: Authors.

Figure A2: Graphical illustration of employment-related intervention effects. Example (ii): Training program for unemployed youth

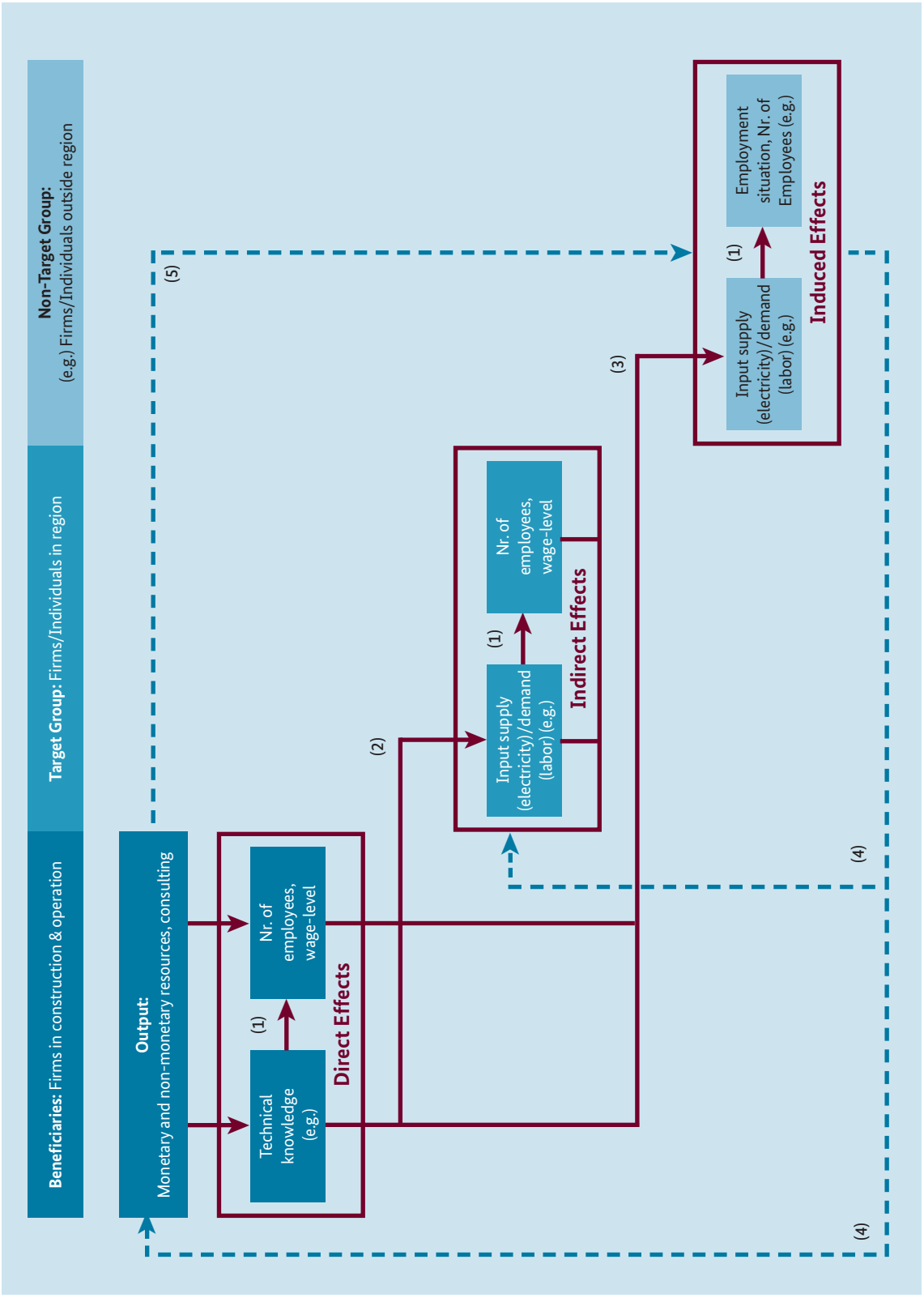


(1) Mediators: (Relative) Employability,... (2) Indirect Effects: Knowledge spillovers, substitution,... (3) Induced Effects: Business environment effects (e.g. supply of qualified labor), multiplier effects (consumption), displacement,...(4) Feedback loops: Knowledge spillovers, labor demand,... (5) Externalities

Source: Authors.

Back to Contents

Figure A3: Graphical illustration of employment-related intervention effects. Example (iii): Construction of a regional power plant



(1) Mediators: Productivity, Competitiveness, ... (2) Indirect Effects: Knowledge spillovers, substitution, multiplier effects, business environment effects,...
(3) Induced Effects: Knowledge spillovers, business environment effects (input supply), multiplier effects (consumption), displacement,...
(4) Feedback loops: Business environment effects, multiplier effects,...(5) Externalities.

Source: Authors.

Back to Contents

Back to Contents

Published by
Deutsche Gesellschaft für
Internationale Zusammenarbeit (GIZ) GmbH

Registered offices
Bonn and Eschborn
Germany

Sector Project Employment Promotion in Development Cooperation

Friedrich-Ebert-Allee 40
53113 Bonn
Germany
Tel. +49 (0) 228 44 60 - 0
Fax +49 (0) 228 44 60 – 1766

Dag-Hammarskjöld-Weg 1-5
65760 Eschborn
Germany
Tel. +49 (0) 61 96 79 - 0
Fax +49 (0) 61 96 79 - 1115

employment@giz.de
www.giz.de/expertise/html/4458.html

Design and layout
Andreas Hesse, Berlin

Printed by
TopKopie, Frankfurt am Main
Printed on FSC-certified paper

Photo credits
Cover: fgl-werketage.de

As of
December 2014

GIZ is responsible for the content of this publication.

On behalf of
Federal Ministry for Economic Cooperation and Development (BMZ)
Division Refugees; migration and employment; reintegration

Addresses of the BMZ offices

BMZ Bonn
Dahlmannstraße 4
53113 Bonn
Germany
Tel. +49 (0) 228 99 535 - 0
Fax +49 (0) 228 99 535 - 3500

BMZ Berlin (Europahaus)
Stresemannstraße 94
10963 Berlin
Germany
Tel. +49 (0) 30 18 535 - 0
Fax +49 (0) 30 18 535 - 2501

poststelle@bmz.bund.de
www.bmz.de