Yearly monitoring of outcomes in supported companies for credible estimates of impact

PRIME POLICY BRIEF #3

August 2014   **www.primepartnership.nl**

Assessing Net-effects of business development support

One key goal of the PRIME project is to identify the impact of CBI and PUM activities on firm-level outcomes, immediate & intermediate outcomes (e.g. knowledge, skills & practices) and ultimate outcomes (e.g. turnover, profit, employment). One of the main challenges in estimating these effects is to establish what would have happened without the support from CBI or PUM: in other words, to assess the net effects that can be attributed to the support. To discard the influence of other factors, we need a design that compares between companies that have and have not received support. In this policy brief, we explain the logic behind our approach in assessing these net effects and, using the M&E data of all supported firms, discard alternative explanations of the changes in performance of the supported companies.

The most conventional design used to measure causal impact is to find a comparison group with similar characteristics that has not received support. Ideally, this would consist of firms that are similar both in observable and in unobservable characteristics. However, in private sector development – with a diversity of sectors, markets and business strategies – it is impossible to find truly comparable subjects. Therefore, we need a process to correct for the bias due to this inevitable 'imperfect match'. Differences in terms of observables (e.g. size, experience, etc.) can be controlled for ex-post using regression analysis, propensity score matching (PSM) or a combination of both. However, controlling for unobservable bias (such as motivation and entrepreneurial behaviour) is more challenging. Moreover, these unobservable characteristics are expected to be quite important in explaining the trend in performance of the firm, even without CBI or PUM support. Random selection of firms from a larger group of eligible firms would be a way to remedy this bias, but proved impossible. The idea of random assignment goes against the rationale and mandate of CBI and PUM to target support to SMEs that have the most potential to create development impact. The (self-)selection of firms receiving support is a reality.

The Centre for the Promotion of Imports from Developing Countries (CBI), the PUM Netherlands Senior Experts (PUM), the Agricultural Economics Research Institute (LEI Wageningen UR) and the Erasmus School of Economics (ESE) develop a joint programme to pioneer impact evaluation methods of support to small and medium enterprises in developing countries. The PRIME programme is designed to develop an approach to data collection that enables the outcomes of PUM and CBI to be tracked while at the same time giving both organisations more insight into their interventions and the opportunity to learn about how they can manage for greater impact.

Cohort design

Reviewing several alternatives, we opted to use a cohort design (shown below). In this design, we collect time-series data from the companies that are granted CBI or PUM support in different years (cohorts). We compare the average status of firms that have already received CBI and/or PUM support with the baseline status of the firms that have not received it yet. For example, in 2015 we can compare cohort one (indicators on performance two year after start of support) with cohort two (indicators at one year after start of support) and cohort three (indicators at baseline). This makes it possible to estimate net effects after one year, after two years and three years. This methodology will give reliable estimates of programme net-effects only if the companies in the cohorts are comparable. Therefore, from the data about the situation of firms before the support started, we check differences in observed pre-programme characteristics. For instance, in the example given above, we will test the similarity of the observed characteristics for cohorts one, two, and three by using baseline and recall data from 2013. If those characteristics prove to be very different, we will use a propensity score matching to get cohorts with similar observable characteristic. We expect that firms asking for PUM-expert advice in 2015 will be quite comparable – in terms of unobservable characteristics such as motivation – to firms that ask for support in 2016, but we foresee that CBI client cohorts will differ much more.

YEAR	cohort 1 (2013)	cohort 2 (2014)	cohort 3 (2015)	cohort 4 (2016)	cohort 5 (2017)
2011	recall	-	-	-	-
2012	recall	recall	-	-	-
2013	baseline	recall	recall	-	-
2014	follow-up	baseline	recall	recall	-
2015	follow-up	follow-up	baseline	recall	recall
2016	follow-up	follow-up	follow-up	baseline	recall
2017	-	follow-up	follow-up	follow-up	baseline

Time-series data

To reduce this threat to validity, next to the above-described counterfactual design (the comparison between cohorts), the data-collection will derive time-series data on the indicators of each firm. These time-series give an estimate of the development of the firm before and after the support started. Observed differences in outcomes through time in each firm cannot be directly attributed to PUM and CBI support. Other "exogenous" variables, economic and political circumstances, for example, can influence issues such as firm practices or profits. Nevertheless, it gives food for thought. The yearly information on the changes in indicators in (groups of) firms will enable 'real-time monitoring'.

CONTROLLING THE MAIN VALIDITY THREATS TO THE DESIGN
The validity of both design methods depends upon the various cohorts or supported and non-supported companies being similar in terms of observable and unobservable characteristics. The observable characteristics will be taken into account in selecting respondents, designing the survey, and matching during data analysis. Unobservable characteristics are by definition more challenging to control, but will be captured partly by the cohort design. Repeating data-collection with different recall periods allows us to detect and control bias.
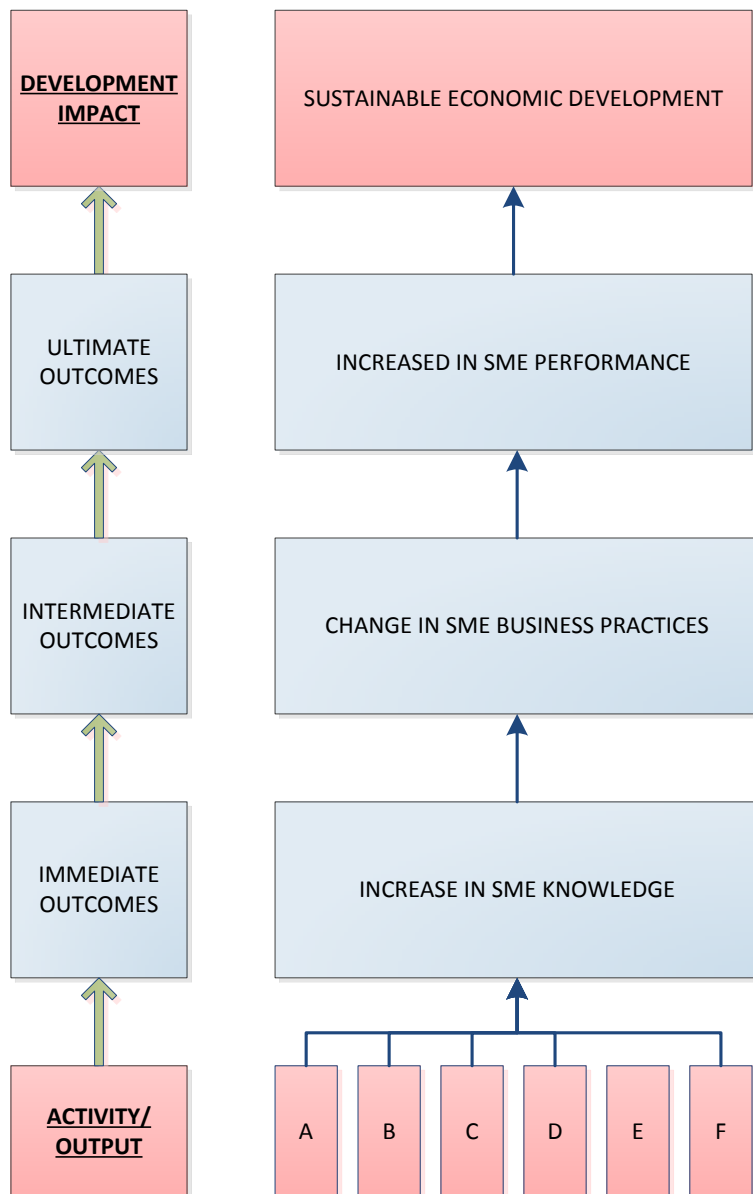
This combination of two quasi-experimental methods – the analysis of time-series and trends in firm clients ('before-after' support) and the comparison of cohorts ('with-without' support) – makes it possible to reflect on two different estimates of effects. This makes the design more resilient to eventual problems in its operationalisation..

MULTIPLE ESTIMATES OF THE EFFECTS OF THE SUPPORT BASED ON THE M&E DATA:

1) Before-after analysis of the indicators in the supported firms
2) Comparison between comparable cohorts of supported firms and companies that are not (yet) supported to estimate yearly effects
3) Heterogeneity analysis of the average effects for different types of firms and different types of support

Selection bias

The cohort design is crucially dependent on comparability of cohorts. The analysis of time-series will result in weaker evidence on impact but is more straightforward. Because it is anticipated that finding comparable yearly cohorts for CBI supported firms will be challenging, the latter will be especially useful for evaluating that support. CBI support is usually within a specific sub-sector where a longer-term support programme is started for the different companies at the same time. If new companies are assisted the year after, this is often under a different programme in a totally different sector. To enable the cohort design and time-series trend analysis, we rely on so-called recall data. Respondents will have to recall their situation. This is not problematic when this concerns the previous year, but might be more challenging when recalling information earlier in time. In our design, the respondents need to recall up to three years before they started receiving support. In the case that respondents over or under estimate data due to different recall periods, this

DEVELOPMENT IMPACT	SUSTAINABLE ECONOMIC DEVELOPMENT
ULTIMATE OUTCOMES	INCREASED IN SME PERFORMANCE
INTERMEDIATE OUTCOMES	CHANGE IN SME BUSINESS PRACTICES
IMMEDIATE OUTCOMES	INCREASE IN SME KNOWLEDGE
ACTIVITY/ OUTPUT	A B C D E F

may induce a bias. We will test for the presence of this type of structural recall bias by asking the same information twice, in two different years. However, we think that these differences will level out when computing group averages. In the cohort design, we observe only the firms that apply to and are granted CIB and PUM support; therefore, we are able to estimate the average programme impact on treated firms. This estimate of impact on only selected firms has policy relevance; it is the best predictor of future impact on firms that have similar characteristics with the firms that were supported in the past.

Indicators

We will determine the impact of both programmes based on the expected 'theory of change'. This means we should at least capture the outcomes at various stages of the theory of change. Based on the information provided by the outputs of the support activities for each type of stakeholder, we will define a typology of support modalities. We will link these to

- immediate outcomes (knowledge)
- intermediate outcomes (business practices)
- ultimate outcomes (firm performance).

Intermediate outcomes are less context specific than immediate outcomes and ideally have generic characteristics that enable benchmarking. Ultimate outcome indicators (performance of the firm's business strategies, e.g. profit, employment, etc.) are more standardized, but are often outside the span of direct influence of a support activity.

> **HARMONISATION**
>
> To maximize the room for comparison with other support programmes, it is essential to harmonize proxy indicators and outcome areas with other studies. To do so, PRIME has adopted various indicators that align with IRIS, an online catalogue of performance metrics. IRIS indicators are well-accepted internationally and are used by relevant partners such as ASPEN, FMO, DCED and Oxfam Novib. Finally, from a practical point of view, we try to keep the proxy indicators in the large N within the limits that allow respondents to provide the information in short interviews and web-based surveys. In addition, we will build as much as possible on already established data-collection procedures in CBI and PUM, and fine-tune and upgrade these systems so that they are integral parts of their M&E systems.

Activities/outputs

In this domain, we include outcome areas that will be used for the identification of the respondent and the characterization of the SMEs, and to explore whether there are different outcomes for different types of SMEs. We will use this data to identify and control for differences between the SMEs that receive support and those that have not.

Outcome areas

We distinguish between immediate and intermediate outcomes on the one hand, and ultimate outcomes on the other. Immediate and intermediate outcomes refer, respectively, to changes in knowledge and practices within the firm, whereas ultimate outcomes refer to the subsequent effect on firm performance – typically captured by turnover, profit and/or employment. These outcomes capture the effects – the underlying theory of change – along

OUTCOME AREAS
1. Finance
2. Management
3. Marketing
4. Sales
5. Human resource management
6. Production technologies
7. Production quality
8. Production efficiency
9. Environmental effect
10. Research and development

the result chain: the advice provided to the firms initially results in improved knowledge and skills, which in turn leads to improved firm performance. The immediate and intermediate outcome areas are categorized into seven different areas mainly based on the structure of existing data collection by CBI .

By collecting data on immediate and intermediate outcomes, and not only on ultimate outcomes related to firm performance, we open a black box that enables us to better understand why the activities result in outcomes for some, but not for others.

Proxy indicators for immediate outcomes - changes in knowledge

During the literature review (see PRIME Policy Brief #1), we reviewed the existing studies on private sector development to learn how they measured outcomes. Many existing tools (e.g. balanced score cards) are very sector specific, meant for internal management only, or requiring extensive questionnaires (such as the audit performed by CBI). We use Likert-scale questions to measure perceived knowledge in each business cluster. We will look at the relative change in the scores, next to their absolute values. In addition, we will ask if they

self-assess whether, in their opinion, the indicator has changed in the previous two years, and to which extent they consider this related to the support provided by PUM or CBI. This helps us to focus our analysis and estimate the span of direct influence of the CBI and PUM support, and point out areas where the changes in indicators result from other factors and actors. Linking these perceived subjective effects to more objectively measureable effects captured by the yearly surveys, we can triangulate the findings to get stronger inferences on impact.

TIME-SERIES
Data will be collected with at least four different time frames:
- ***Pre-project data***: refers to the situation before the support of PUM or CBI started. Ideally, we get three years of recall data on these indicators to allow comparison of the trends in different cohorts.
- ***Project data***: refers to the period after the start of CBI and PUM support. We will capture the data on a yearly basis to allow us to see the trend and to make the inter-cohort comparisons.
- ***Post-project data***: refers to the period after the end of CBI and PUM support. Within the current programme budget, we propose to measure outcome indicators at least one year after the support has ended, though, preferably we will continue data collection for a longer time period to capture longer term effects.

Proxy indicators for intermediate outcomes - changes in business practices

Tools used by other support organisations to register changes in business practices tend to be quite context specific, and, generally, they require extensive questionnaires. Because of the broad contents of practices, we ask them to rate themselves in relation to competitors or similar firms and to indicate if this relative ranking has changed in the last year. Next to these subjective questions, we included more objective questions on practices. When possible, we follow the standard IRIS indicators. In deciding on these indicators, we had to

balance between including sufficient generic indicators and keeping the indicators to a minimum.

Proxy indicators for ultimate outcomes – changes in firm performance
We propose a set of questions to measure firm performance structured around the following categories: (i) employment, (ii) export, and (iii) financial performance. In the literature on development and business economics, these classes are the most widely used. Employment is disaggregated by gender. This corresponds to the focus on gender awareness in general, and of CBI and PUM in specific.

Proxy-indicators for development impact - sustainable economic development
We need information that allows a reflection on the contributory role of the ultimate outcomes on development. We will monitor various indirect indicators such as employment and average wages. Direct attribution to the support interventions of CBI and PUM is impossible at this level because a change in these indicators is far outside the span of direct influence of CBI and PUM. In order to verify the broader contribution of the programme's intermediate and ultimate outcomes on sustainable economic development, we will use focus group discussions and interviews with key informants in the case study countries. In these case studies, we can combine the outcomes of the quantitative analysis with secondary data about sector dynamics to 'reason' the contributory role of increased exports and product upgrading by firms.

Data collection
In the current M&E system of both CBI and PUM, experts involved in the support activities provide most of the data. Within the context of PRIME, we will also collect data directly from companies. This allows us to verify the quality of the data provided by both type of informants and control for potential biases as well as under or over reporting by the firm or by experts.

Additionality
A key assumption behind the support provided by PUM and CBI is that this support is not available yet offered by experts that reside in the development country. To verify this key assumption, insight should be gathered on the question of whether or not the firms could have accessed equivalent support themselves, either by self-financing or from other (commercial) actors. Therefore, we introduced some questions about the presence and use of other service providers.

Authors: Giel Ton[a] , Fédes van Rijn a, Job Harms[b], Karen Maas[b], Haki Pamuk[c]

This policy brief is an output from research funded by CBI & PUM Netherlands. The views and opinions expressed in this document are those of the authors and do not necessarily reflect the views of CBI & PUM Netherlands..

A=LEI, Wageningen University
B=Erasmus School of Economics, Erasmus University
C=Wageningen University