

IDS Bulletin

Transforming Development Knowledge

Volume 53 | Number 1 | February 2022

THEORY-BASED EVALUATION OF INCLUSIVE BUSINESS PROGRAMMES

Issue Editors **Giel Ton and Sietze Vellema**



Notes on Contributors	iii
Introduction: Contribution, Causality, Context, and Contingency when Evaluating Inclusive Business Programmes Giel Ton and Sietze Vellema	1
Systems, Sapiens, and Systemic Change in Markets: The Adopt-Adapt-Expand-Respond Framework Ben Taylor and Jake Lomax	21
Using Theory-Based Evaluation to Evaluate Systemic Change in a Market Systems Programme in Nepal Edward Hedley and Gordon Freer	43
Assessing the Contribution to Market System Change of the Private Enterprise Programme Ethiopia Giel Ton, Ben Taylor and Andrew Koleros	63
The Search for Real-Time Impact Monitoring for Private Sector Support Programmes Fédes van Rijn, Haki Pamuk, Just Dengerink and Giel Ton	87
Monitoring Systemic Change in Inclusive Agribusiness Sietze Vellema, Greetje Schouten and Marijn Faling	103
Assessing Contributions Collaboratively: Using Process Tracing to Capture Crowding In Marijn Faling	123
Understanding Behaviour Change in Theory-Based Evaluation of Market Systems Development Programmes Jodie Thorpe	141
Glossary	165

The Search for Real-Time Impact Monitoring for Private Sector Support Programmes*

Fédes van Rijn,¹ Haki Pamuk,² Just Dengerink³ and Giel Ton⁴

Abstract An increasing consensus exists in the impact evaluation literature on using detailed theory-based evaluations to evaluate complex programmes such as private sector development (PSD) programmes. At the same time, PSD managers expect periodic and timely (so-called 'real-time') input from evaluators to improve programmes throughout their implementation. This article presents insights from real-time theory-based monitoring and evaluation shaped by the needs of policymakers in two Dutch PSD programmes. To learn about their experiences, we held in-depth interviews with researchers and policymakers involved in the evaluation. The interviews indicated that theory-based evaluation improved reporting on the programmes' contribution to higher-level impact areas and credibly quantified the importance of that contribution. The insights showed too that real-time monitoring and evaluation of PSD programmes requires more flexibility in data collection and increased interaction with mid-management.

Keywords private sector support, business coaching, economic development, theory-based evaluation, impact analysis.

1 Introduction

Private sector development (PSD) programmes aim to contribute to overall economic development through providing business support services (such as technical assistance, management provision, export training) or financial aid (Schulpen and Gibbon 2002). Over the past two decades, pressures on development budgets have increased the demand to show results at the impact level in order to legitimise public funding. While many PSD programmes have monitoring systems in place that collect information on outputs (for example, number of companies trained) and to a lesser degree on uptake (such as use of

training), they face difficulties when asked to report on the impact on business performance. PSD programmes, therefore, need to upgrade their monitoring system in such a way that they can respond to donors' expectations with regard to evaluation. The Donor Committee for Enterprise Development (DCED) recommends building a results management system that tracks, among other things, the effects of support activities as attributable changes to business performance, job creation, and export performance (DCED 2017).

To get quantitative estimates of impact that can be attributed to programme activities, most impact evaluation approaches (Khandker, Koolwal and Samad 2010) ask for the comparison of programme outcomes before and after the programme, and ideally between supported and unsupported firms. In the case of PSD programmes, impact evaluations usually require information on business performance indicators such as employment, sales, and exports before and after participation in the programme, and information to correct for contextual influences for a regression-based estimate. The programmes contribute to changes in high-level indicators such as employment, sales, and exports of firms, but these net-effect indicators of impact are not very actionable and useful for adaptive management during programme implementation, because at best, they are only available after some time (Apgar, Hernandez and Ton 2020). Often the programme management learns about the impact only when the decisions about continuation, adaptation, or finalisation of the PSD support have already been made.

Therefore, not many PSD programmes choose to rely on quantitative research designs for computing attributable net effects. Instead, most PSD programmes provide illustrative examples of their relevance at impact level in a more qualitative way, through case studies. However, case studies rarely provide a representative picture of the quality and impact of the portfolio of activities. The challenge for evaluators is, thus, to find practical ways to report the size and importance of the support that are lean enough to be incorporated into a programme's monitoring and evaluation (M&E) system for portfolio-level monitoring but are rigorous enough to result in credible estimates of the overall impact of the PSD programme to allow a reflection on its relevance, effectiveness, and efficiency. There is an increasing consensus among the quantitative and qualitative-oriented impact evaluators that for complex programmes – such as PSD programmes – programme theories need to be the backbone of an impact evaluation design (Chen 1994; Blattman 2008; White 2009; Bates and Glennerster 2017; Davey *et al.* 2018). In these theory-based evaluations, the data collection is designed in response to key assumptions in the programme theory.

The policy relevance of impact evaluations depends on the extent to which programme management has access to

these findings to refine and adapt their programmes. Ideally, information for monitoring, evaluation, and learning is shared throughout the implementation in 'real time', using methods and processes that enable adaptive management (Giordano 2017).

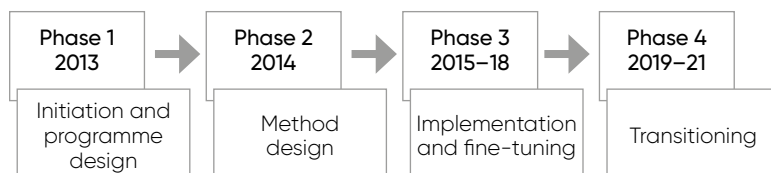
This article reflects on how theory-based mixed-methods impact evaluation can assess the importance and impact of PSD support in terms of offering accountability to the funders, while serving the information needs of programme managers. For this purpose, we distil lessons learned from the implementation of the Pioneering Real-time Impact Monitoring and Evaluation (PRIME) programme between 2013 and 2021. In PRIME, two large Dutch PSD organisations, the Centre for the Promotion of Imports from developing countries (CBI) and the Netherlands Senior Experts (Programma Uitzending Managers, or PUM), used similar tools to report on the impact of their PSD support. The programmes differ in aims but both have business coaching as a common approach. CBI promotes exports from developing countries through sectoral programmes that provide advice, counselling, and export market entry support to small and medium-sized enterprises (SMEs) and business support organisations. PUM organises business-level and sectoral missions that help SMEs to improve their business practices. Both CBI and PUM are funded by Dutch official development aid because they aim to generate additional (export) sales and employment in those countries, and therefore contribute to sustainable and inclusive economic growth.

To review the lessons learned in the implementation of our theory-based mixed-methods impact evaluation approach to assess the impact of PSD support, Section 2 explains the evolution of PRIME between 2013 and 2021. Section 3 reflects on the PUM and CBI programme managers' experiences with the approach and their assessments about its policy relevance. The section builds on information from in-depth interviews held in 2017 and 2021, and a workshop with a wider group of programme stakeholders conducted in 2017. We discuss user feedback and the main trade-offs and tensions that researchers and programme managers encountered in implementing PRIME. Finally, Section 4 provides recommendations for a better integration of theory-based impact evaluation and M&E systems of PSD programmes.

2 PRIME approach

The PRIME partnership was established in 2013 by CBI, PUM, the Erasmus School of Economics (ESE), and Wageningen University and Research (WUR), to develop and implement a methodology to monitor and evaluate the real-time impact of private sector development support by PUM and CBI. We distinguish four phases in PRIME (see Figure 1). In this section we describe each phase.

Figure 1 The four phases of PRIME



Source Authors' own.

2.1 Phase 1: initiation and programme design

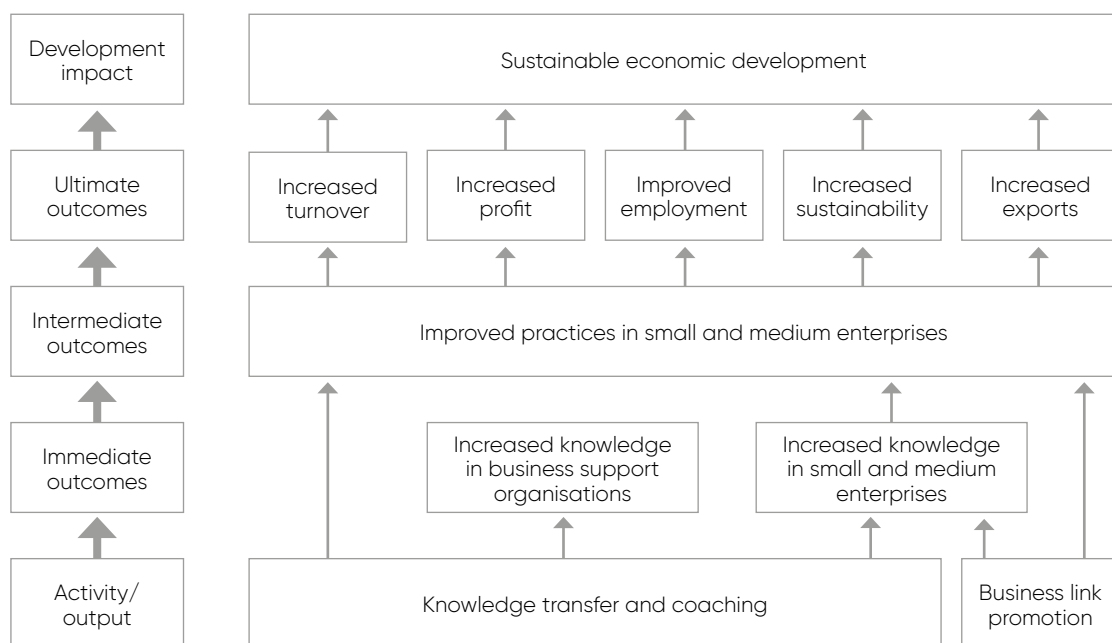
Following an instruction from the Dutch Ministry of Foreign Affairs in 2011 (DGIS 2011), all Dutch PSD organisations with a budget above €10m were made responsible for evaluating the impact of their work on sustainable development outcomes. The guidance also emphasised the need to show net effects of impact and the use of counterfactual research designs to do so. Many of these organisations struggled with this need and started to experiment with methods to generate credible evidence.

The idea for the PRIME partnership emerged in 2012. It was the fruit of informal discussions during a series of seminars hosted by WUR for the 'PSD Platform', where most Dutch PSD support organisations are represented. There were three reasons for establishing the PRIME partnership. First, the necessity of reporting the impact of private-sector support on the harmonised impact indicators defined by the DCED (jobs, revenues, and scale) – in other words, **accountability needs**. Second, the difficulty of going beyond 'before/after' measurements and the use of comparison groups – **a methodological need**. Third, a desire for meaningful impact evidence which can be used during the implementation of programmes – **a learning need**.

CBI and PUM, both prominent members of the Dutch PSD Platform, decided to address the challenges together as their organisations had complementary objectives. They also aimed to work more closely together and were accountable to the same governmental body and civil servants. The assumption was that a better understanding of each other's strengths, using a similar method for benchmarking effectiveness, would create synergies between both programmes. They approached Wageningen University and Research (WUR) and ESE to help them. WUR had a track record in developing evaluation methods for value chains in agriculture, forestry, fisheries, natural resources, and consumer markets. ESE had a track record in performance measurement and development of corporate social responsibility programmes of companies.

The organisational structure of the PRIME programme was designed to ensure the involvement and ownership of CBI and PUM, while at the same time maintaining sufficient independence

Figure 2 Simplified intervention logic of SME support provided by CBI and PUM



Source Authors' own.

to meet the quality criteria for external evaluations defined by the Dutch Evaluation Office (IOB). PRIME had a Programme Board, consisting of the managing directors from CBI, PUM, and higher management in WUR and ESE, plus an Advisory Committee, consisting of six external representatives, including the Ministry of Foreign Affairs, the IOB, the International Trade Centre (ITC), the International Initiative for Impact Evaluation (3ie), and two other knowledge institutes (Panteia and The Hague University of Applied Sciences).

During this inception phase, overarching theory of change (ToC) charts (using the term 'intervention logics') were developed to allow theory-based evaluation and to sketch the preliminary mix of core method. The research partners facilitated the process, as the concept of a ToC chart was not yet used by either organisation at the time. Several underlying assumptions for the change process were added to the chart, including identification of risks and plausible unintended effects, for the main causal links in the chart (Ruyter de Wildt *et al.* 2013).

The overarching intervention logic (Figure 2) was used to identify indicators at several outcome levels that could capture the effects of CBI's export promotion and PUM's business coaching. The ToC chart used the disaggregated outcome categories as suggested by Mayne (2001): immediate outcomes (knowledge);

intermediate outcomes (business practices); and ultimate outcomes (firm performance).

The programme design document (Ruyter de Wildt *et al.* 2013) proposed a mix of core methods plus additional methods and routines to anticipate the main validity threats to these (Ton 2012). The core methods were: (1) a literature review; (2) a cohort design to collect panel data; and (3) case studies in six countries to explore whether the support led to systemic change. The programme design was approved by the end of 2013, leading to the next phase of PRIME.

2.2 Phase 2: method design

In phase 2, the method of PRIME was operationalised. Most attention was on the development of indicators for the intermediate outcomes, as these were deemed to be less context-specific than the immediate outcomes and, therefore, have more generic characteristics that enable benchmarking. Ultimate outcome indicators (firm performance in terms of profit, employment, exports, and so on) are even more standardised, but, as anticipated by the researchers in PRIME, could be outside the span of influence in terms of time or programme effect (Ton, Vellema and Ge 2014).

Literature review. A key activity in this phase was a literature review on the current evidence regarding SME support (Harms, Ton and Maas 2014), which benefited from several extensive systematic reviews that became available at the time (Grimm and Paffhausen 2014; Piza *et al.* 2016). The literature review confirmed the plausibility of the ToC but indicated a lack of evidence about the assumed employment effects of PSD support at the sector and national level. The study also showed a list of indicators used by other scholars to track changes of intermediate and ultimate outcomes.

Survey design. PUM had already collected information on immediate outcomes, while CBI collected data mostly on the ultimate outcomes. Data on intermediate outcomes (business practices) were hardly collected. The indicators had to capture meaningful change and be general enough to be relevant for different types of firms that operate in various economic sectors. The immediate and intermediate outcome areas were related to seven different areas of business management as distinguished by CBI in their company auditing system.

To complete these data, we proposed an online survey to collect data, not only on ultimate outcome indicators (sales, exports, employment), but especially on the intermediate outcomes on knowledge and practices with a combination of self-assessment questions and observable business management practices. The self-assessment questions would give real-time feedback and the observable business management practices would help to

triangulate and validate these perceptions of impact over time. The yearly online survey asked the firm managers to assess the extent to which PSD support contributed to changes in the firm's business management. The survey was tested and adapted to be around 15 to 20 minutes, using the Qualtrics platform.

Case studies. Parallel to this, we designed the case studies in five countries. Considering budget constraints and following advice of the Advisory Committee – 'better do one thing good than two things flimsy' – we refrained from collecting additional data from the supported firms by subcontracted data collection firms, but focused on semi-structured interviews with beneficiaries, experts, and other key stakeholders.

2.3 Phase 3: implementation

Between 2015 and 2018, the PRIME evaluation methods were implemented. On the one hand, this meant collecting and analysing data; on the other hand, this meant regular interaction between researchers and staff from CBI and PUM. The research team had at least monthly interaction with the Consultation Group, quarterly face-to-face meetings with the Advisory Committee, and bi-annual meetings with the Programme Board. PRIME produced monitoring reports after each survey round or country visit, quarterly newsletters for wider audiences, and yearly research briefs for CBI and PUM. Reporting on such data on a regular basis was new for both organisations, and in the case of PUM, led to a special section in their yearly reports.

For PUM, out of 5,353 firms that were invited to take part in one or more surveys, 2,779 completed them. Similarly, for CBI, the online survey was sent to all firms that had received support from CBI in the previous three years. The number of firms that responded to this online survey was 318 in 2014; 369 in 2016; and 348 in 2017. Overall, the response rates – between 30 per cent and 52 per cent – were considered good for online surveys and were largely thanks to the intensive follow-up by PUM and CBI staff.

The qualitative case studies reflected the diversity of the sectors and economic conditions in which PUM and CBI operate. They made use of interviews with beneficiaries, experts, and other key stakeholders in the different sectors and countries. In doing so, the case studies helped to illustrate the programme's effects in terms of types of business knowledge and practices that changed, using insights from the survey. More importantly, it helped to identify enablers and barriers of effectiveness of the support modalities and how the support related to other sector-level innovation processes in the country.

After the first survey, there was a need to adjust the methodology. There was an idea to compare the performance during the programme with the trends three years before the support had started; however, this proved over-optimistic. It became clear

that PUM and CBI did not manage to generate these 'three-years-before' data on the key performance indicators (ultimate outcomes), and that the survey could not fill the gap due to fatigue and recall bias in these estimates.

Moreover, in the inception phase, we assumed that the cohorts of firms that were selected by PUM and CBI would be 'on average' similar. For PUM this assumption proved plausible (van Rijn *et al.* 2018a). However, for CBI it became clear that the number and type of supported firms depended on the sector and countries that were prioritised in each four-year period, which made the inter-cohort comparisons of indicators unreliable (van Rijn *et al.* 2018b). The solution found was to use a pooled regression with the time of participation of a firm as a covariate. This indicator captures the effect of the time after the firm has had the first contact with the support programme.

Additional to the online survey, CBI decided to collect some key performance metrics through support staff in-country. The so-called 'certified results' efforts managed to collect rich data about exports of the firms during a session where CBI staff looked at the financial reports together with the firm owner; the export data were presented in a reporting format signed off by both. The decision to conduct this resource-intensive data collection was not coordinated with the PRIME researchers who had argued in the design phase that this type of data could be too far out of the sphere of direct influence of CBI, and that capturing data on changes in business practices was more important from a resource-efficiency standpoint. In the end, however, the certified results data on exports were highly valuable for the final impact evaluation.

2.4 Phase 4: transitioning

After the funding for PRIME ended, in 2018, the collaboration between WUR and PUM continued, though at a lower intensity and within a different scope. The core objective was to continue to provide independent impact monitoring based on online surveys with PUM's M&E system. The case studies were discontinued as PUM did not receive sufficient insights on impact from these resource-intensive studies.

WUR continued with their support to extract insights from PUM's data by applying econometric techniques. Data collection with the online survey and quality assurance of these data shifted entirely to PUM. This resulted in a yearly expanding and increasingly rich time-series about impact. The increase in data points available enabled the research team to better estimate the lagged effects of the business coaching on SME performance. The key deliverable was to provide PUM with the externally validated ('certified') outcome and impact estimate in PUM's Annual Report, in some way similar to the accountant statements in the financial report.

CBI decided not to continue with PRIME. Reasons were related to the investments in terms of time and money, compared with the perceived benefits of this additional data collection to the existing efforts and the certified results exercises. However, during a 2021 interview with CBI, one interviewee did indicate that, as an indirect effect of PRIME, the organisation became more aware of the importance of good technical systems with good indicators, and that this had led to significant investment to improve these systems. Although CBI did not use the tools developed in PRIME, they strengthened their data systems in response to the experience with PRIME.

Moreover, the perceived benefit of the PRIME programme – namely, assessing the causal effect on exports – became less evident after 2018, as the accountability requirements for PSD had changed. The new guidelines for results reporting indicate that net effects do not always need to be shown (DGIS-RVO 2017); when a PSD programme could show a significant contribution to a complex change process, it was allowed to report the total change generated without estimating the attributable part of this total change. CBI, with activities in a specific sector and with firms, sector organisations, and governments, could show their impact more easily than PUM, and the 'certified results' became sufficient to report their contribution.

3 User feedback

To get a clearer picture of the experiences with the real-time monitoring and evaluation in the PRIME partnership, various in-depth interviews were held with researchers and policymakers involved in the programme. In total, 19 interviews were held in September 2017, with three follow-up interviews in March and April 2021. Moreover, in September 2017, a joint workshop was held with 15 participants who had been involved in phases three and four, to reflect on the use of the information for management decisions.

The interviews indicated that the PRIME partnership had helped both CBI and PUM to increase their accountability to donors by creating trust in their strengthened M&E systems. The partnership was positively appreciated by the Ministry of Foreign Affairs, which helped to secure continued donor funding. A staff member of PUM said: 'We have opened our organisation for a bunch of scientists. It was a bit of a gamble for us, you don't know in advance what comes out of it.'¹⁵ Respondents indicated that support from researchers was and remains essential in designing new questions and analysing the data. Both CBI and PUM indicated that shifting data collection to the implementing organisation, without external researcher involvement, might also negatively impact on the credibility of the results.

The workshop and interviews also gave insight into user feedback in relation to the use of the real-time evaluation approach for

learning: (1) helping implementing organisations to become more impact-oriented; (2) using results from the real-time evaluation in management decisions; and (3) balancing associated workload for implementing staff with the learning from the real-time data.

3.1 Making organisations more impact-oriented

According to respondents, the real-time evaluation approach of the PRIME partnership has made their respective organisations more impact-oriented: 'PRIME has brought more focus on regular impact monitoring. The focus is now more on the quality of the mission, rather than the number of missions.'⁶

The interaction with researchers helped both organisations to sharpen their intervention logic and better define the different impact pathways of their organisations and associated outcome areas. PRIME helped them to broaden their perspective on measuring results and to go beyond the traditional focus on monitoring outputs by including more, and more appropriate, indicators at the intermediate outcome level: 'Now there is much more focus on intermediate outcomes. In each project we now report on intermediate outcomes.'⁷

The academic perspective of the researchers helped the staff of both organisations to think more critically about what to measure and how to organise data collection and data management in a way that assists them in producing portfolio-level reporting. As one respondent of an implementing organisation put it: 'Due to PRIME, we have invested much more in ICT [information and communication technology] and the role of our organisation in doing data collection and analysis.'⁸

3.2 Use of impact data in management decisions

It is clear from the interviews that findings from PRIME were used to shape discussions on the future direction of PUM and CBI activities. A respondent from CBI indicated:

Many of the conclusions of the PRIME study were integrated in our latest five-year strategy. In line with the conclusions of PRIME, the strategy proposed to move beyond merely European markets, adopt more digital ways of working, pay attention to gender and youth and identify larger companies to work with.⁹

Another example comes from PUM. In 2016, the external evaluation of PUM activities (van der Windt *et al.* 2016) used PRIME data on changes in business practices to suggest that PUM should shift its portfolio more to micro and small businesses. However, PUM used the PRIME data on business performance to argue that while impact on the business practices of micro and small companies was indeed higher, in terms of employment and turnover, the effect of PUM was higher on larger companies. The discussion shows that the PRIME data were useful for

strategic decision-making and enabled a better reflection on effectiveness.

However, as areas for improvement, CBI and PUM indicated that the communication of results by the team of researchers was often too technical and complex, and the results did not relate directly enough to the day-to-day practices of the organisations to influence more operational decision-making. It was suggested that more visually attractive and more simply written research products in an early phase of implementation could have been helpful. As one of the implementing organisations put it in the evaluation workshop in 2017: '[In our organisation] you need to present your material on a serving platter, in an attractive and accessible way for the results to be used.'¹⁰ Based on this feedback, the research team in 2018 dedicated more attention to the visual layout and readability of the final reports, and included a separate chapter with recommendations to the management in the subsequent reports.

Moreover, several respondents felt that the PRIME partnership could have embedded the researchers in the offices of CBI and PUM. More personal interaction could have increased the degree to which PRIME data and analysis were used by people in both organisations. 'For a partnership to work, you need to see each other regularly. You need regular discussions with programme managers for it to come to life.'¹¹ At the same time, one respondent stressed that true **ownership is also required**: 'People will not take it seriously and/or use PRIME. This is especially true now because PRIME was "invented" by people that are not working in the organisation any more'.¹²

Another interviewee from PUM said that these moments of interaction were also important for them as M&E officers to help them to become more visible in their own organisation: 'I need PRIME to connect with the rest of the organisation for it to receive support. This means connecting to people from knowledge management, management accounting and business development.'¹³ Other respondents also indicated that more frequent sharing of results on both sides would have improved the level of engagement and learning in the partnership. One suggestion was taken up in phase four with PUM, where at the start of each year a meeting was organised to identify certain strategic themes for which the online survey could be used to collect additional information. As a result, additional topics such as gender, food security, and indirect effects were integrated into the last versions of the online survey.

3.3 Reasonable balance between workload and learning

The above two points sketch out the benefits of PRIME. These benefits need to outweigh the costs, especially when the learning, and not the accountability to the donor, becomes the main goal. Both the online survey and the case studies needed support

from staff. Several respondents from CBI and PUM indicated that the implementation of the PRIME research activities was too heavy a burden for some of their colleagues and partners in the field. There were suggestions to reduce the data collection to the online survey only, eliminating the qualitative part of the real-time evaluation. In contrast, other respondents found the qualitative case studies to be the most relevant for their work, while the results of the quantitative survey were felt to be more challenging to interpret and translate into action. Especially in CBI, a feeling emerged that PRIME was complicated and time-consuming, and had insufficient value for managing the different sector and country programmes; this was one of the reasons why CBI did not engage in phase four of the PRIME partnership. Furthermore, it was harder for CBI to translate the research outputs to the day-to-day work practice and decision-making processes. The support is so diverse that it is difficult to learn from average overall trends. Disaggregation of the econometric results was limited due to the relatively small number of firms involved in each sector.

4 Conclusions

This article presents insights from a theory-based impact evaluation of business coaching and export promotion that navigated the needs of the stakeholders in PSD programmes for learning and accountability, and that provided real-time information for adaptive management.

PRIME succeeded in its aim of improving the reporting of the PSD programmes' contribution to higher-level impact areas (export, employment), and of quantifying the importance of this contribution in a credible way, as demanded by the donors at the time. The key elements that made it convincing were the clear charts with appropriate indicators, the yearly time-series, and the sophisticated econometric analysis by the researchers. This contributed to a more convincing programme evaluation and a more informative report for donors.

Aside from this accountability aim, the ambition of PRIME was to improve effectiveness of the programmes by supporting monitoring, evaluation, and learning processes on an ongoing basis and feeding these with regular (real-time) insights during implementation. While the PSD organisations and policymakers benefit from the theoretical perspective and rigour of the theory-based approach, followed by a theory-based evaluation, the data collection on an ongoing basis has a cost in terms of financial and human resources. The accountability requirement to report net effects at impact level provided a clear incentive to invest in more rigorous survey-based evaluation approaches. When the funder's accountability requirements shifted towards a more qualitative approach to assess contribution and additionality, this incentive became less apparent for CBI.

When learning is concentrated at portfolio level, as was the case in PUM, the data collected in the online survey on firm practices and performance need to be aggregable and, therefore, more general. For CBI, the sector managers required more granular data than PRIME provided, and this explains why they considered the associated workload too high. Making sector-focused versions of a survey could have been a golden midway for CBI as well as PUM, including some questions and indicators that are applicable to a certain sector or country only, which could increase its relevance for the staff and experts involved in the support. The online survey modules with perception questions and contribution scores proved a flexible tool for creating sector-specific versions (see Ton, Taylor and Koleros, this *IDS Bulletin*).

Another important lesson learned in PRIME is that the monitoring and evaluation should connect to the PSD implementers' everyday reality – both their work processes and their information needs. For this purpose, it is critical to regularly involve not only the M&E staff, but also mid-management, such as country or sector managers, and communication staff. More frequent encounters or workshops can assist evaluators, researchers, and policymakers to engage more in joint sense-making of the evaluation results. Regularly 'harvesting' the data needs and key questions that PSD programmes are facing may help to ensure a better match between the research analysis and the reality of the organisations, and may improve the ownership of the evaluation.

The common ground of these two strategies – flexibility in data collection and increased interaction with mid-management – is the search to improve the usefulness and timeliness of theory-based evaluations and to find an appropriate balance between accountability and learning. PRIME helped to navigate this search, took steps in the right direction, but also showed that the road to 'real' real-time monitoring and evaluation is still a long one.

Notes

- * We would like to acknowledge the contribution to the design and implementation of the PRIME programme of all the stakeholders involved. Also, we would like to thank Martijn Ramaekers and Andy Wehkamp for their feedback on a draft version of this article.
- 1 Fédes van Rijn, Senior Researcher, Wageningen University and Research, Netherlands.
- 2 Haki Pamuk, Senior Researcher, Wageningen University and Research, Netherlands.
- 3 Just Dengerink, Independent consultant, Food Systems, Netherlands.
- 4 Giel Ton, Research Fellow, Institute of Development Studies, University of Sussex, UK.
- 5 PUM staff, interview, 18 September 2017.
- 6 PUM staff, evaluation workshop, 26 September 2017.
- 7 CBI staff, interview, 31 March 2021.

- 8 PUM staff, interview, 26 September 2017.
- 9 CBI staff, interview, 31 March 2021.
- 10 PUM staff, interview, 26 March 2017.
- 11 WUR staff, interview, 12 March 2017.
- 12 PUM staff, interview, 29 March 2021.
- 13 PUM staff, interview, 18 September 2017.

References

- Apgar, M.; Hernandez, K. and Ton, G. (2020) 'Contribution Analysis for Adaptive Management', *Briefing Note*, London: Overseas Development Institute
- Bates, M.A. and Glennerster, R. (2017) 'The Generalizability Puzzle', *Stanford Social Innovation Review* 15.3: 50–54
- Blattman, C. (2008) 'Impact Evaluation 2.0', presentation to the Department for International Development (DFID), London, UK, 14 February
- Chen, H-T. (1994) *Theory-Driven Evaluations*, London: SAGE
- Davey, C. et al. (2018) *Designing Evaluations to Provide Evidence to Inform Action in New Settings*, CEDIL Inception Paper 2, London: Centre of Excellence for Development Impact and Learning
- DCED (2017) *The DCED Standard for Measuring Achievements in Private Sector Development: Control Points and Compliance Criteria*, London: Donor Committee for Enterprise Development
- DGIS (2011) *Protocol Resultaatsbereiking en Evalueerbaarheid in PSD*, The Hague:
- DGIS–Directie Duurzame Economie (DDE)–Dutch Evaluation Office (IOB)
- DGIS-RVO (2017) *15 Methodological Notes – Instructions for Calculation, Validation and Reporting of Performance Indicators*, The Hague: Dutch Ministry of Foreign Affairs
- Giordano, N. (2017) *Monitoring, Evaluation and Learning: Adaptive Management to Achieve Impact Results*, Care International blog, 18 January
- Grimm, M. and Paffhausen, A.L. (2014) *Interventions for Employment Creation in Micro, Small and Medium-Sized Enterprises in Low- and Middle-Income Countries: A Systematic Review*, Frankfurt: KfW Bankengruppe
- Harms, J.; Ton, G. and Maas, K. (2014) 'Literature Review: The Impact of Advisory Services and Export Promotion on SME Performance', *PRIME Policy Brief 2*, Wageningen: Wageningen University and Research
- Khandker, S.R.; Koolwal, G.B. and Samad, H.A. (2010) *Handbook on Impact Evaluation: Quantitative Methods and Practices*, Washington DC: World Bank
- Mayne, J. (2001) 'Addressing Attribution Through Contribution Analysis: Using Performance Measures Sensibly', *Canadian Journal of Program Evaluation* 16.1: 1–24
- Piza, C. et al. (2016) 'The Impact of Business Support Services for Small and Medium Enterprises on Firm Performance in Low- and Middle-Income Countries: A Systematic Review', *Campbell Systematic Reviews* 12.1: 1–167

- Ruyter de Wildt, M. de; Maas, K.; Ton, G. and Harms, J. (2013) *Pioneering Real-Time Impact Monitoring and Evaluation in Small and Medium Enterprises (PRIME) – Final Report Phase I: Programme Design*, Wageningen: Wageningen University and Research
- Schulpen, L. and Gibbon, P. (2002) 'Private Sector Development: Policies, Practices and Problems', *World Development* 30.1: 1–15
- Ton, G. (2012) 'The Mixing of Methods: A Three-Step Process for Improving Rigour in Impact Evaluations', *Evaluation* 18.1: 5–25
- Ton, G.; Vellema, S. and Ge, L. (2014) '**The Triviality of Measuring Ultimate Outcomes: Acknowledging the Span of Direct Influence**', *IDS Bulletin* 45.6: 37–48 (accessed 18 October 2021)
- van der Windt, N. et al. (2016) *Evaluation of PUM Netherlands Senior Experts 2012–2015: An Independent Evaluation Study Commissioned by the Netherlands Ministry of Foreign Affairs*, Rotterdam: Erasmus University Rotterdam
- van Rijn, F. et al. (2018a) *Verification of PUM's Intervention Logic: Insights from the PRIME Toolbox*, The Hague: Wageningen Economic Research
- van Rijn, F. et al. (2018b) *Verification of CBI's Intervention Logic: Insights from the PRIME Toolbox*, The Hague: Wageningen Economic Research
- White, H. (2009) 'Theory-Based Impact Evaluation: Principles and Practice', *Journal of Development Effectiveness* 1.3: 271–84

This page is intentionally left blank