



Article

Evaluating the impact of business coaching programmes by taking perceptions seriously

Evaluation

2023, Vol. 29(1) 73–90

© The Author(s) 2022



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/13563890221137611

journals.sagepub.com/home/evi**Giel Ton** 

Institute of Development Studies, University of Sussex, UK

Fédes van Rijn

Wageningen Economic Research, The Netherlands

Haki Pamuk

Wageningen Economic Research, The Netherlands

Abstract

The paper addresses the challenges of evaluating the impact of business coaching programmes with a varied portfolio of firms working across sectors and countries. Observable indicators of changes in business management practices are rarely relevant across sectors. Therefore, evaluators need to rely on the perceptions of the managers who have received coaching. We designed an online survey to compare the effectiveness of business coaching within a portfolio and across programmes. The survey was applied to the portfolio of two private sector development programmes. We derived so-called ‘contribution scores’ from individuals’ perceptions of how business management practices had changed and their perceptions of the role of business coaching in bringing about these changes. The survey included some features to reflect on response reliability. We show that the tool seems fairly reliable for comparative analysis and helped to identify the types of firms and contexts where business coaching support appears more effective.

Keywords

business management, contribution analysis, contribution scores, impact evaluation, self-assessment

Introduction

This article presents a method to help reflect on the relevance and effectiveness of development support. We illustrate the use of the method in the evaluations of two business coaching

Corresponding author:

Giel Ton, Institute of Development Studies, University of Sussex, Library Road, Brighton BN1 9RE, UK.

Email: g.ton@ids.ac.uk

programmes targeting small- and medium-sized enterprises (SMEs) in developing countries. Both programmes were funded by the Netherlands Directorate-General of International Cooperation (DGIS). We used a (bi)annual online survey of SME managers receiving coaching as the core data collection tool to evaluate whether the business support programme could be considered a significant contributor to improved business performance.

The paper is organised as follows. First, we present the background of the intervention modality – business coaching – and explain why we opted for a perception-based survey tool. Second, we describe the business coaching activities of the two organisations involved, the theory of change, and the background to the impact evaluation. Third, we describe the perception survey and explain how to derive ‘contribution scores’ from it, highlighting the design elements that allowed us to reflect on the reliability of the data. Fourth, we present the results, illustrating how contribution scores can help assess whether the intervention was a significant contributing factor in the wider, complex configuration of factors that created the development outcomes. Finally, we reflect on how the method could be used in other impact evaluations.

Business coaching

Business coaching consists of a collaborative relationship between a business manager and a professional coach, aiming to strategically develop a successful business. It entails a relationship

in which the coachee and the coach collaborate to assess and understand the coachee and his or her leadership tasks, to challenge current constraints while exploring new possibilities, and to ensure accountability and support for reaching goals and sustaining development. (Blackman et al., 2016: 460; Ting and Hart, 2004: 116)

Business management and business coaching are complex processes. The business coach tries to influence and improve the behaviour of managers of SMEs who are constantly making decisions about changes in business strategy based on their interpretation of a set of complex, interrelated dynamics, including market signals, technical restrictions, the legal environment, human resource capacity, plans of competitors, and shifting financial realities. The business practices and management behaviour typically reported in the management literature include “more innovation and creativity, higher levels of initiatives or more flexibility and adaptability” (Blackman et al., 2016: 470). The coaching–coachee relationship is complex due to its sensitivity to personal characteristics and contextual features. The effects of the business coaching are in the manager’s head and may not be directly reflected in business performance indicators, as these are contingent upon many other factors and actors within and around the SME that shape the business strategy. Estimating the precise net effects of business coaching support on organisational outcomes is, therefore, virtually impossible (Blackman et al., 2016). And, even if it were possible, the estimates would not be particularly useful in informing decisions on implementation, replication and scaling: “business coaching . . . is not going away any time soon even if the evidence that it works is not yet (or will ever be) strong enough. It is part of the human nature to invest in hope rather than solid outcomes” (Athanasopoulou and Dopson, 2018: 85–86).

When evaluating the impact of business coaching on business performance, one must acknowledge that the support provided to managers is, at most, an INUS condition – ‘an insufficient but non-redundant part of a condition which is itself unnecessary but sufficient for the

result' (Mackie, 1974: 62). The first part of this INUS definition is clear: the support needs to be combined with other conditions in order to work. The second part of the definition emphasises the need to consider alternative causal explanations – factors that may explain the result but have nothing to do with the support. Evaluating an intervention as an INUS condition follows the generative logic of causal inference (Stern et al., 2012). Data for generative inference can be collected using a wide range of methods (Pawson, 2013), including econometric methods that have a sequential (rather than generative) logic of pattern detection (Aston et al., 2021; Pawson, 2013). The important point is that the data allow a better understanding of the context sensitivity of the coaching outcomes (Athanasopoulou and Dopson, 2018: 84).

Private sector development is an integral part of the international development strategy of many governments and development agencies. This support involves the transfer of public funds to private ventures, which is only legitimate when it helps to attain public goals. Therefore, private sector development programmes that are publicly funded are under increasing pressure to rigorously monitor and evaluate the results of their work (DGIS, 2011; Donor Committee for Enterprise Development (DCED), 2017; Organisation for Economic Co-operation and Development (OECD), 2016).

Typical policy instruments for private sector development programmes include technical and financial support to SMEs and indirect technical assistance through sector-based organisations and specialised government programmes. The rationale for supporting SMEs is based on the assumption that they have more growth potential, are more financially constrained, and have less access to advisory services than larger firms. Inclusive business programmes target specific SMEs that have the potential to benefit the poorer strata of the population, either as producers or consumers (Ton and Vellema, 2022).

Recent systematic reviews of business and leadership coaching (Athanasopoulou and Dopson, 2018; Blackman et al., 2016; Pandolfi, 2020) cover studies that mainly concern coaching in developed countries and large organisations. Their findings are helpful when reflecting on evaluation research designs. Blackman et al. (2016: 467) point to the limitations in the design of studies that only use subjective participant evaluations and self-reported subjective evaluations of effectiveness, which may partly explain why 'all of the reviewed studies reported that the target coaching programme was seen by most participants as effective in some way' (p. 469).

The positive effects of training and business coaching are measured and reported in many different ways. Some studies computed productivity growth and return on investment (ROI) by comparing before and after measurements in task productivity, but ignored the selection bias due to working with SMEs that were already eager to innovate and grow (Blackman et al., 2016; De Meuse et al., 2009; Humphrey and Navas-Alemán, 2009). These ROI estimates were often based on very specific and peculiar indicators that limit comparability (Grant, 2012). In more recent studies, changes in behaviour (Lawrence and Whyte, 2014), coaching satisfaction scores (Robson and Bennett, 2000) and perceived improvements in self-efficacy (Crompton et al., 2012) are the preferred indicators to assess and compare the effectiveness of coaching and training.

The theoretically possible negative effects mentioned in the literature – for example, that time spent with the coach could have been better used, or the negative effect of coaching-induced goal-setting (as suggested by Athanasopoulou and Dopson, 2018) – cannot be completely discarded, but are very unlikely to be important in the real-world practice of business coaching to SMEs. Regarding the opportunity costs of the time spent with the coach, the

coachee can limit these by simply cancelling meetings should they feel that the relationship is not providing the expected benefits. Regarding the risk that a coachee starts to implement process innovations or technologies that later prove inappropriate or ineffective, the supposed negative causal link might underestimate the vast experience and reflective capacity of the manager concerned. SME managers who use business coaching are highly likely to use information from many more sources, and not just the coach, before making important business decisions.

Private sector development programmes that offer business coaching to SMEs in developing countries have a target group of firms that are eligible to receive their support. They cannot rely – as is usual in developed countries – on a market mechanism where the coachee pays for their services; public funds are involved, and public goals need to be reached, which means it is often necessary to target firms that are more likely to respond positively to the support offered. Unfortunately, however, most studies on business coaching do not provide much detail about the type of firms in which coachees work. Few studies focus on the characteristics of the coachee and their organisational context; most studies focus on the characteristics of the coach (often in view of further professionalisation) and the quality of the relationship between the coach and the coachee (Blackman et al., 2016). This is unfortunate because the information about firm-specific conditions could be used for portfolio management and policy discussions about targeting business coaching support to specific types of firms or countries.

Design challenges

In private sector programmes, theory-based evaluations are becoming the dominant way to learn about and assess the effectiveness of support modalities (Ton and Vellema, 2022) due to the severe real-world constraints to quasi-experimental approaches for causal inference of net effects in complex programmes (Lemire et al., 2019; Mackie, 1974; Stern et al., 2012). The outcomes of business coaching typically result from configurations of multiple conditions that are often quite specific to each firm in a sector. Some types of firms might respond better to business coaching than others. Moreover, the support modalities are not standard and one-off but evolve according to the needs of each firm; therefore, the quality of implementation of business coaching varies enormously from firm to firm and from expert to expert.

Moreover, more mundane reasons preclude the mainstream quasi-experimental designs for evaluating support to formal SMEs. Business support programmes that target SMEs often need to focus on specific sectors where the number of firms may be too low to find a meaningful comparison group. Random assignment of the treatment (business coaching) is rarely possible because most programmes deliberately want to work with firms that have a greater chance of success or are part of specific company networks. Theoretically, a regression discontinuity design – making use of oversubscription and a measurable selection threshold – could be a good option. Unfortunately, in the real world, both conditions are often not met. Oversubscription of firms that self-select for support is rare in practice. And (self-) selection is highly dependent on personal networks, such as membership of industry associations or chambers of commerce, rather than being related to quantitative measures of eligibility or proposal quality with a clear cut-off point.

Performance outcomes are also at the limits of the sphere of influence. Blackman et al. (2016) argue that managers' perceptions about the benefits of business coaching may not necessarily relate to the SME's performance; similarly, business performance indicators alone

may not reflect the effectiveness of business coaching. It is also the case that firms not receiving support have little incentive to disclose commercially sensitive business and performance information. Given these challenges, there is a need for evaluation methods that can assess the effectiveness of business coaching support without collecting data from a comparison group.

A common way to gain information on the outcomes of business coaching for business management practices that are comparable across different sectors is to ask coachees directly. Many real-world impact evaluations collect coachees' perceptions, sometimes through a survey and almost always through qualitative interviews. However, perception data are often considered the weakest form of evidence and is commonly ignored in systematic reviews because of the high risk of bias. We argue, instead, that perception data collected in a survey with some design elements that allow reflection on the reliability of responses can be very useful and informative. We acknowledge the limitations of perception data for strong inferences about causal effects while showcasing its strengths for comparing outcome patterns across intervention modalities and as a way to test the plausibility of a contribution claim.

Illustrative cases: CBI and PUM

The two interventions we have used as examples in this paper have business coaching as a core component of their support, targeting SMEs in developing countries. SMEs differ substantially from micro-enterprises because of the size of their workforce. In these two cases, the SMEs are all formally registered and have a workforce of between 30 and 65 permanent staff.

Business coaching programmes

Our evaluation focuses on the impact of business coaching support provided by two publicly funded private sector development programmes: the Centre for the Promotion of Imports from developing countries (CBI) and PUM Netherlands senior experts (PUM). Both programmes partnered in the Pioneering Real-time Monitoring and Evaluation (PRIME) research programme together with Wageningen University and the Rotterdam School of Management.

The CBI, established in 1971, is an agency of the Netherlands Ministry of Foreign Affairs. It aims to strengthen the sustainability of SMEs in developing countries. It works across 47 countries and 27 sectors, targeting SMEs that want to export products to European markets. It offers a structured training programme to improve SMEs' export capacity through various activities such as on-site coaching, training in Europe, participation in international trade exhibitions, and business-to-business linkages. The support provided to each firm typically lasts for 4 years. On average, CBI manages a yearly portfolio of between 700 and 800 firms. Its business coaches have varied backgrounds, though most are consultants who specialise in the sector.

PUM was established in 1978 by the Netherlands employers' federation VNO-NCW, with financial assistance from the Ministry of Foreign Affairs. Its mission is to boost performance in the SME sector in developing countries by mobilising the wealth of knowledge amassed by business managers at the end of their careers. PUM has around 2000 experts and a network of about 150 local representatives in 30 countries who help identify SMEs interested in receiving its support. The SME pays for the expert's local travel, accommodation and interpreter fees, and PUM pays for their flight, insurance and visa costs. In 2014, PUM organised 1900 missions to 71 countries by 1283 experts (van der Windt et al., 2016). All coaches have previous management experience in a similar sector, generally in a developed country.

Research partnership

The Netherlands Ministry of Foreign Affairs (DGIS, 2011) requires that all publicly funded private sector development organisations evaluate the impact of their interventions. The Ministry asks for credible research designs to assess: (1) effects on job creation, (2) business profits, and (3) reach. These three impact indicators followed the recommendations by the DCED to harmonise impact indicators that allow a comparison of the effectiveness and cost-efficiency of different private sector development support programmes. Initially, DGIS asked for net-effect estimates using a comparison group design. However, a follow-up methodological guideline by the Directorate-General for International Cooperation-Netherlands Enterprise Agency (DGIS-RVO, 2017) allowed a ‘contribution approach’ for reporting the impact of private sector support. It asked organisations to report the results of the firms they work with while demonstrating that the intervention was a significant contributory factor in improving business performance. ‘An intervention is significant if one can reasonably expect and hold the project responsible for achieving progress toward significant changes in behaviour of the entrepreneur or other positive outcomes for workers, based on the scope of provided support’ (DGIS-RVO, 2017: 4).

In 2013, CBI and PUM approached Wageningen University and Research (WUR) and Erasmus University for a research partnership to develop and implement a system of data collection that could satisfy these accountability requirements. The approach adopted by this research partnership was to develop an interlinked research design to verify key steps in the intervention logic using a mix of quantitative and qualitative methods, following the logic of theory-based evaluation and reporting as evidence-based contribution stories. Following the example of Ton (2012), for each method, the validity threats to the anticipated type of inferences were explored and, where possible, addressed by including additional research methods. The mix of methods for data collection included administrative data from each programme’s monitoring system, qualitative case studies in six countries using stakeholder interviews, and an online survey undertaken by all supported firms several times during the period they were receiving support (van Rijn et al., 2018a, 2018b). In this paper, we focus on the online survey component, which captured the perceptions of change among business owners supported by the programmes, including the design elements that helped reflect on the reliability of the data, which was the main validity threat to our conclusions.

The survey module included statements with self-assessment questions that helped us to compute indicators of the effectiveness of the business coaching support for eight areas of business management. These indicators – contribution scores – combine the Likert-type-scale answers to two questions: one that asks about perceived improvements in business management practices, and another that asks about the perceived importance of the coaching activities. Subsequently, the eight scores are averaged into an overall score and used in lagged regressions to test whether the perceived effects on practices correlated positively with the firm’s economic performance 1 year later.

Evaluation approach

The evaluation used a theory-based evaluation approach, verifying whether the intended activities took place and to what extent the expected causal links in this theory were supported by evidence (Mayne, 2011, 2019b). We began by articulating the basic intervention logic (see Figure 1) and main impact pathways, depicted as a sequence of causal steps in the theory of

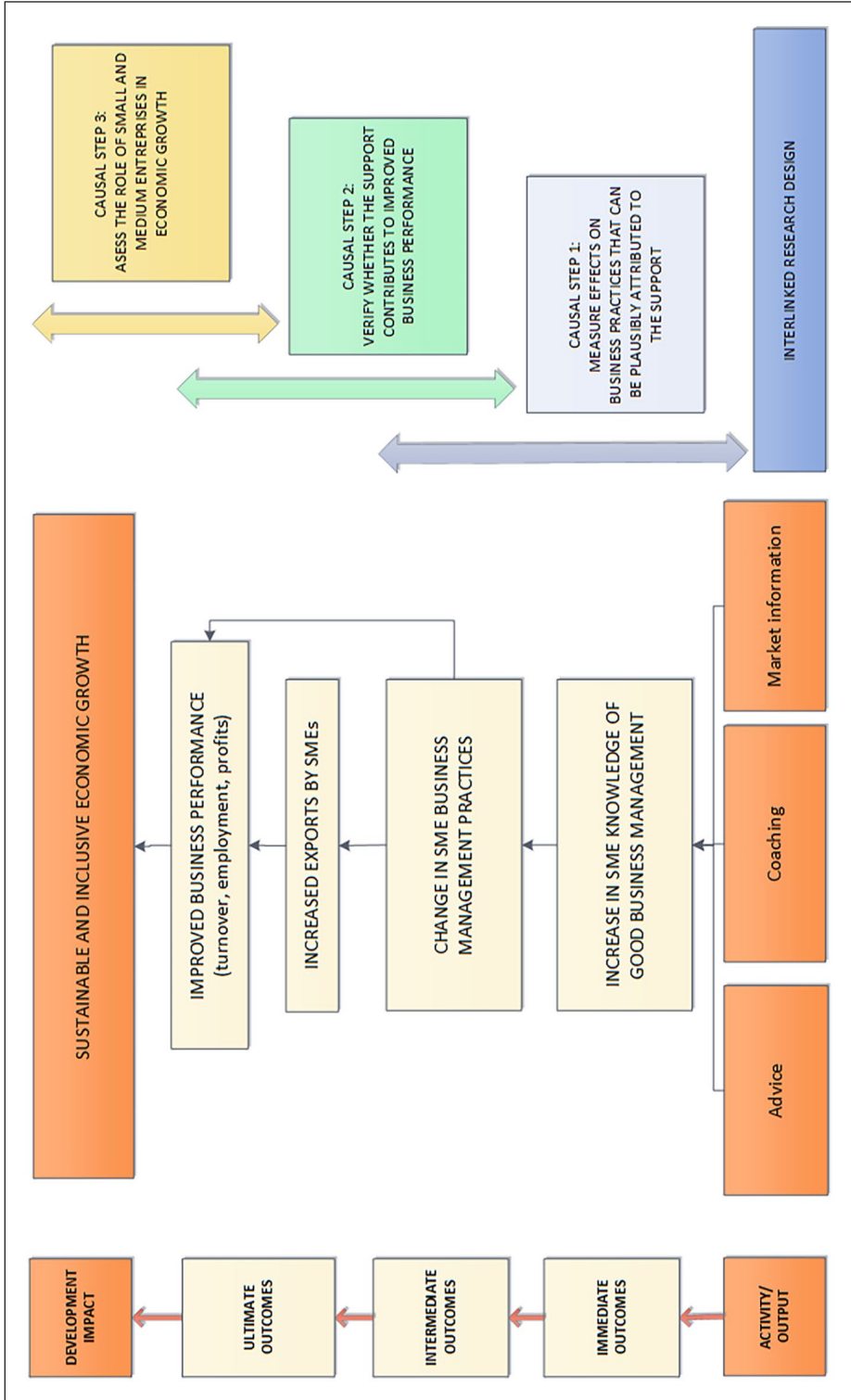


Figure 1. Basic intervention logic of the Centre for the Promotion of Imports from developing countries (CB) and PUM Netherlands senior experts (PUM).

change (Funnell and Rogers, 2011; Mayne, 2001; White, 2009). The figure shows two impact pathways: one with direct coaching support to SMEs and the other with indirect support to firms via business support organisations (BSOs).

The impact evaluation focused on changes in firms' business management practices and whether these changes contributed to improved business performance.

We opted for an interlinked research design with three analytical moments. The first analytical moment was based on the data from the (bi)annual online survey that captured managers' perceptions about whether their firm's management practices had changed and what influence the support from CBI or PUM has had on this change process. The second analytical moment was using an econometric regression to assess whether this change in business management practices had a positive effect on the firm's economic performance: namely, the increase in the firm's profits or exports, based on data available in the monitoring systems of both organisations. The third analytical moment was the verification of the causal step in the theory of change between the improved performance of SMEs and macro-level sustainable economic development, through a literature review and interviews with business managers and key experts in six qualitative case studies in specific sectors in six countries (reported in van Rijn et al., 2018a, 2018b).

Perceptions and contribution scores

Fortunately, at evaluation design, much of the data on each firm's characteristics and economic performance was available in the respective programme's existing monitoring systems. Both CBI and PUM already collected administrative data on the firms' performance at intake and continued to monitor it in subsequent years.

The crucial information that was missing related to the changes in the firm's business management practices in response to the business coaching provided. To capture this information, we developed a short (15–20 minutes) online survey with questions appropriate to many different types of firms and economic sectors, which could be used as part of regular monitoring. Therefore, we developed a set of questions to assess the perceived changes resulting from coaching support in eight areas of business management, as used by CBI in its company audits assessing the export potential of the SME.

For each of the eight areas, there were three sub-questions. First, we asked the respondent to assess their firm's business management practices compared to competitors in their sector. Second, we asked them to assess any change in this area over the past 12 months. Third, we asked them to assess the influence of PUM or CBI activities in relation to this change. All three questions had Likert-type scale answers and included 'don't know' as an option (see Table 1).

While the answers to each of these questions separately already give useful information, we decided to go a step further and convert two of the three self-perception questions into a new variable that would indicate the importance of business coaching in improving business management practices. We called these variables 'contribution scores' (CS). Table 2 shows the ranking we used to convert the Likert-type scale answers on both questions into a contribution score. All answers which stated that the support provided by CBI or PUM had had no perceived influence were coded as having a CS of zero. No survey respondent reported a negative change in business practices due to PUM or CBI. All remaining combinations were ranked in logical order and converted to percentage-like scores, with rank 8 being the highest (CS of 100%) and rank 0 being the lowest (CS of 0). With these eight contribution scores,

Table 1. Elements in the survey module.

Outcome areas	Perception questions	Likert scale answers
1. Financial management	Question 1	Much worse (1), worse (2), same (3), better (4), much better (5)
2. Leading, planning and organising the business	How do you assess your company's practices compared to others in the sector?	
3. Marketing techniques to increase sales		
4. Quality requirements of (inter)national buyers	Question 2	Strong decrease (-2), decrease (-1, no change (0), increase(1), strong increase (2)
5. Ways to retain, motivate and train employees	How have your company's practices in this area changed over the past 12 months?	
6. Efficient ways of organising the production process or service delivery		
7. Effects of the business on the environment	Question 3	No effect (0), very little (1), some (2), quite a bit (3), a lot (4)
8. Ideas about new products and services	Has [CBI/PUM] influenced this change?	

Source: Authors' own.

CBI: Centre for the Promotion of Imports from development countries; PUM: PUM Netherlands senior experts.

Table 2. Conversion into contribution score.

Answer to question 2: How have your company's practices in this area changed over the past 12 months?	Answer to question 3: Has [CBI/PUM] influenced this change?	Contribution rank (0–8)	Contribution score (%)
Strong decrease	No effect	0	0
Decrease	No effect	0	0
No change	No effect	0	0
Increase	No effect	0	0
Strong increase	No effect	0	0
Increase	Very little	1	13
Strong increase	Very little	2	25
Increase	Some	3	38
Strong increase	Some	4	50
Increase	Quite a bit	5	63
Strong increase	Quite a bit	6	75
Increase	A lot	7	88
Strong increase	A lot	8	100

Source: Authors' own. No respondent indicated a decrease in business practices where they perceived CBI or PUM to have influenced that change. Therefore, these theoretically possible combinations are not considered in the table. CBI: Centre for the Promotion of Imports from developing countries. PUM: PUM Netherlands senior experts.

each covering a different area of business management, we computed a simple, unweighted overall average contribution score for each firm.

Reliability checks

The use of self-assessment and perception questions created obvious validity threats. Therefore, we introduced some features into the survey that helped to reflect on the reliability of the data collected. We did this in three ways.

First, we applied the same survey multiple times in both coaching programmes. This helped to reflect on the consistency of outcome patterns related to the different coaching modalities in both organisations. We expected the contribution scores to vary between the eight areas of business management, with scores for each area remaining relatively similar over time.

Second, we randomised the sequence of the eight business management practices for which the perception questions were asked, using the features of the online survey software (Qualtrics). This design element was expected to reduce potential fatigue bias, whereby respondents might lose focus or seriousness in responding to the online survey questions. If fatigue bias were an issue, at least all eight areas of business management practices would be affected similarly. Therefore, the pattern of scores across those eight areas would still be indicative of meaningful differences. We applied the list randomisation in all survey rounds to maximise longitudinal data quality. The downside was that this made it impossible to empirically demonstrate the effectiveness of this design feature by comparing the patterns of survey data collected with and without list randomisation.

Third, Likert-type scale answer categories assume that respondents are willing and able to make relatively fine-grained self-assessments. We added a design element that would help us reflect on this critical assumption. Apart from the question, *How do you assess your company's practices compared to others in the sector?*, we asked respondents about 10 observable business management practices that are generally considered to be associated with good business management (see Table 3). Many practices are specific to the type of SMEs supported by CBI and PUM and derived from a literature review by Harms et al. (2014) during the inception phase of the evaluation, with the most relevant studies being de Mel et al. (2012), Robson and Bennett (2000), Drexler et al. (2014) and Field et al. (2010). These observable practices, for which presence or absence was recorded, are compared with the answers to the question that asked how the firm's practices compared with similar firms in the sector. We expected a positive correlation between the self-assessed status in each area and the presence of the corresponding observable practice(s) in the firm.

Results

Assessing survey data quality

The online survey was administered three times during the research. For PUM, the surveys were administered in 2015, 2016 and 2017. The survey invitation was sent to all firms that had been visited by an expert up to 3 years before. Out of 5353 firms that were invited to take part in one or more surveys, 2779 completed them (a response rate of 52%). Similarly, for CBI, the online survey was sent to all firms that had received support in the previous 3 years. The number of firms that responded was 318 in 2014, 369 in 2016 and 348 in 2017 – a response rate of 35, 52 and 40 per cent, respectively.

The pattern of the contribution scores across each of the eight areas proved to be remarkably consistent over time for both programmes (see Figures 2 and 3). And more interestingly, it shows a meaningful difference between CBI and PUM in the area(s) for which their coaching proved relatively more or less effective.

For almost all areas of business management, the correlation coefficient of the association between perceived status in practices compared with other firms in the sector and the observed proxy indicators was positive and statistically significant (see Table 3). Using the

Table 3. Correlation estimates between self-assessed status and observable business management practices.

Question 1: How do you assess your company's knowledge/practices compared to other companies in your sector? Much worse (1), same (2), better (3), better (4), much better (5)	Presence of observable business practice All dummy variables are coded 1 when the practice is observed	CBI 2016	CBI 2017	PUM 2015	PUM 2016	PUM 2017
Financial management	Having specialised software or financial statements verified by control outside the company	0.05 [258]	0.23*** [188]	0.17*** [540]	0.05 [654]	0.06* [960]
Leading, planning, and organising the business	Having a marketing plan	0.08 [257]	-0.07 [248]	0.10** [541]	0.14*** [655]	0.19*** [968]
Marketing techniques to increase sales of your product or service	Having promotion materials	0.08 [265]	0.10 [253]	0.17*** [533]	0.19*** [657]	0.14*** [976]
Quality requirements of (inter)national buyers	Systems to learn about clients' opinions on its products and services	0.06 [259]	0.20*** [238]	0.22*** [523]	0.19*** [621]	0.10*** [928]
Ways to retain, motivate and train employees	All employees have a contract	0.03 [258]	0.01 [193]	0.09** [557]	0.10** [679]	0.06* [1027]
	Policies in place to monitor and ensure workers' safety	0.07 [243]	0.14 [230]	0.07 [490]	0.05 [607]	0.05 [919]
Efficient ways of organising the production process or service delivery	A documented quality assurance system	.06 [249]	0.12* [237]	0.08* [523]	0.14*** [620]	0.08** [909]
Effects of the business on the environment	A system to monitor effects on the environment	0.18*** [258]	0.04 [250]	0.23*** [494]	0.14*** [637]	0.08** [966]
Ideas about new products and services	Introduction of new products	0.16** [260]	0.26*** [251]	- [577]	0.30*** [664]	0.30*** [1005]
	Introduction of new processes	0.19*** [255]	0.29*** [250]	- [577]	0.13*** [647]	0.22*** [979]
Average of the self assessment of status	Average of the observable practices	0.14** [269]	0.27*** [256]	0.38*** [577]	0.26*** [726]	0.25*** [1085]

Source: Original data.

The numbers in brackets show the number of observations used in the correlation analysis (Spearman). Averages are computed for all respondents. Not all respondents provided information on all eight areas or ten practices. CBI: Centre for the Promotion of Imports from developing countries. PUM: PUM Netherlands senior experts. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. - indicates that the question was not part of the specific survey round.

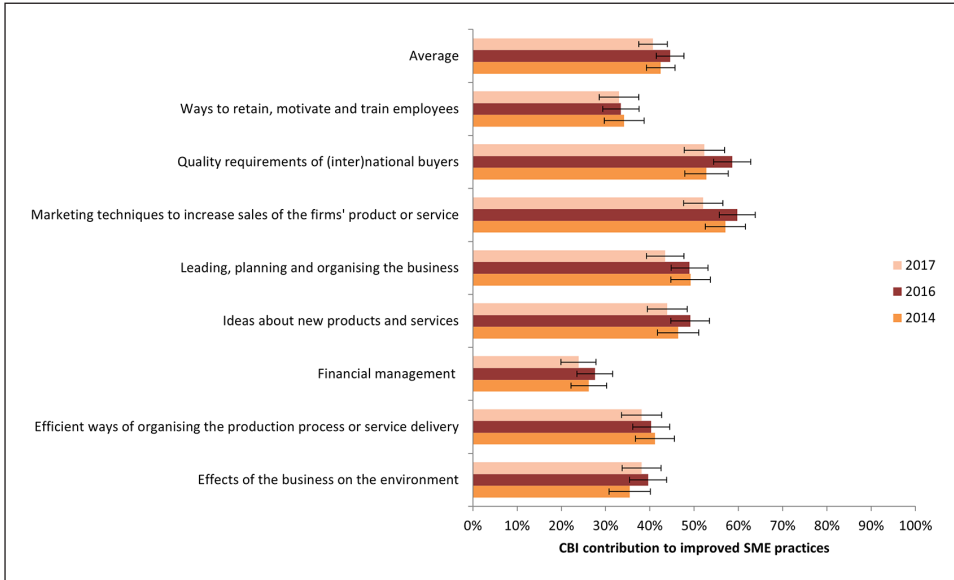


Figure 2. Contribution scores for the Centre for the Promotion of Imports from Developing Countries (CBI) in the period 2014–2017.

The bar figures indicate the contribution scores for CBI and the black lines show the confidence intervals measured at 95 per cent confidence levels.



Figure 3. Contribution Scores for PUM Netherlands senior experts in the period 2015–2017.

The bar figures indicate the contribution scores for PUM and the black lines show the confidence intervals measured at 95 per cent confidence levels.

interpretation of Funder and Ozer (2019), most of the Pearson correlations show a very small effect ($0.05 \leq r < 0.10$) or a small effect ($0.10 \leq r < 0.20$). Considering the heterogeneity of the firms and the multiple other factors that influenced them to adopt certain business practices, the small correlation coefficient was not surprising. The consistent

positive coefficient gave us the confidence that the answers to the self-perception questions were likely to be an imperfect but good-enough reflection of reality.

Verifying contribution to business performance

The similarity of the survey questions about business management practices for firms supported by CBI and PUM enabled us to present comparative graphs to inform discussions by both organisations about the relative strengths of their support. This comparison (benchmarking) reinforced the awareness of potential synergies within the management of both organisations and between them. The latter was especially informative for the Dutch Directorate-General for International Cooperation. For example, Figure 2 shows that CBI had the highest contribution scores in the domains of marketing techniques and quality requirements of international buyers. Figure 3 shows that PUM seems relatively more effective in improving ideas about new products and efficient ways of organising the production process. Both CBI and PUM only marginally influence the business management practices related to financial management. Overall, CBI has higher average contribution scores than PUM, which may reflect the longer duration of its business coaching support compared to PUM. As explained earlier, CBI supports firms over 4 years to start or expand their exports to countries in the European Union (EU), while most PUM experts limit their support to a one-off mission of 2 weeks.

We used the average contribution score in a lagged regression that explored whether this contribution to changed business management practices influenced each firm's performance. The regression specification differs between PUM and CBI because their support modalities are somewhat different, and the preferred business performance indicator used also differed, with exports to the EU as the dependent variable for CBI.

As an illustration, and to reduce the number of tables in this article, we present only the specification and results of the regression for CBI. The regressions tests whether there is a relationship between the contribution of CBI programmes to business management practices and the growth of the firm's exports to the EU

$$\Delta Exports_{i,2016} = \beta_0 + \beta_1 CS_{i,2014} + \theta_i + \phi_s + \gamma_i + \delta_c + \epsilon_{i,2016} \quad (1)$$

where i shows the firms. Again, θ_i , ϕ_s , α_t , γ_i and δ_c are business size, sector, year, cohort and country fixed effects (FEs) respectively. $\Delta Exports_{i,2016}$ is the growth in the firm's value of exports (in logarithms) between 2015 and 2016. $CS_{i,2014}$ is the average contribution score, as perceived in 2014. To increase the plausibility that business coaching is a significant contributing factor to increased exports to the EU, we hypothesise that the coefficient estimate for β_1 is positive and statistically significant.

Table 4 reports the results of the estimation of the regression model (1). We find that the contribution of CBI's business coaching support to business management practices is indeed associated with export growth while controlling for business size, sector, cohort and country FEs. We acknowledge that this causal effect may still result from reverse causality and selection bias – the coaching may be to firms that grow most, 'picking the gazelles' (Humphrey and Navas-Alemán, 2009). Nevertheless, the findings increase our confidence in the causal assumption in the intervention logic that business coaching contributes to the export performance of firms who received support.

Table 4. Regression results for the relationship between contribution scores for CBI firms and growth in exports to EU countries.

Growth in the firm's value of exports to EU countries (in log) between 2015 and 2016	Models			
	(1)	(2)	(3)	(4)
Average contribution score (2014)	0.008**	0.008**	0.008**	0.011**
Constant	-0.285*	-0.185	-0.457	-2.388***
Observations	70	70	70	70
R ²	0.05	0.06	0.13	0.58
Estimation method	OLS	OLS	OLS	OLS
Business size FE		Yes	Yes	Yes
Sector FE		Yes	Yes	Yes
Cohort FE			Yes	Yes
Country FE				Yes

Source: Original data.

The dependent variable is winsorised at 5 per cent level to exclude outliers. Robust standard errors clustered at the country level are reported in parentheses. CBI: Centre for the Promotion of Imports from developing countries;

OLS: ordinary least squares; FE: fixed effect.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Exploring enablers of effectiveness

Moreover, the regular monitoring data and the online survey responses could be used to explore contextual enablers or barriers to the effectiveness of business coaching. We could explore for which countries, sectors and firm size the support appears to be more effective, as measured by each firm's average contribution score. We estimated the following model

$$CS_{i,t} = \theta_i + \phi_s + \pi_c + \omega_c + \alpha_t + \gamma_i + \epsilon_{it}. \quad (2)$$

The variables in regression model (2) are similar to those used in model (1). Therefore, we do not repeat the explanation of the symbols used.

Because our index variable for contribution score is bounded between 0 and 100, we estimate a Tobit model. We report robust standard errors clustered at the country level. Table 5 reports the coefficient and standard error estimates for CBI. The average contribution score of CBI is about 23 points lower in Europe and 17 points lower in the Middle East and North Africa (MENA) compared to the Asian countries. The coefficients suggest that, on average, a firm from Asia perceived more effect of the CBI support in their business management practices.

Contrary to the dominant belief in CBI (van Rijn et al., 2022), the analysis showed that firms from lower- and middle-income countries reported lower contribution scores than those from least developed countries. This finding suggests a higher additionality of CBI support in countries with fewer alternative (domestic) business support services to coach firms. A similar analysis for PUM showed that the supported SMEs had a significantly lower contribution score than firms classified as micro-enterprises. This result was surprising because PUM had never aimed to work with micro-enterprises. The inclusion of a substantial number of micro-enterprises in their portfolio (9%) had not been noticed before. The contribution scores proved useful for detecting heterogeneous effects of the business coaching support and informing discussions about targeting (van Rijn et al., 2022).

Table 5. Regression analysis to explore differences in CBI's contribution scores by firm and country characteristics, dependent variable average contribution score of CBI (0–100).

Variables	Categories	Coefficient estimate	Standard error
Business size	Micro (base)–	–	–
	Small	4.66	(3.51)
	Medium	–0.57	(3.33)
Region	Asia (base)	–	–
	Europe	–22.63***	(5.81)
	Latin America	–4.25	(4.22)
	MENA	–17.43***	(5.31)
	Sub-Saharan Africa	3.27	(4.51)
Country	Least developed (base)	–	–
Income group	Lower-middle income	–10.57***	(3.97)
	Upper-middle income	–13.64***	(4.52)
Sector	Agricultural, fishery & forestry (base)	–	–
	Consumer products	6.49	(4.66)
	Industrial products	–5.54	(5.04)
	Services	6.24	(4.15)
Constant		0.42***	(0.05)
Observations		597	

Source: Original data from Pioneering Real-time Monitoring and Evaluation (PRIME).

The table reports the estimates from a regression analysis where the dependent variable is the average contribution score for CBI. The explanatory variables are listed in the second column. We also control for year and cohort-fixed effects in the regression. All explanatory variables are dummy variables. We use the Tobit estimation and report relative marginal effects at mean levels compared with the reference category. Robust standard errors clustered at the country level are in parentheses. CBI: Centre for the Promotion of Imports from developing countries. MENA: Middle East and North America.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$, – indicates reference category.

Discussion

We have presented details of our attempt to answer the challenging question of whether business coaching can be considered a significant, non-redundant contributory factor for improving business performance, using relatively simple methods to verify three causal steps in the intervention logic.

We argue that contribution scores provide a lean, flexible and compelling way to assess the impact of development interventions. The survey module with perception questions can be integrated into most survey-based impact evaluation designs. We introduced design elements that helped to reflect on the quality of the survey data. Of course, passing these checks does not provide conclusive evidence of reliability; however, they substantively increased our confidence that the data reflected reality to a fair degree.

The significant association between the average contribution score of each firm and their export growth, controlling for other influencing factors, allowed us to increase our confidence that the business coaching of PUM and CBI indeed could be considered a significant contributing factor to improved business performance. Moreover, the explorative multivariate analysis detected contextual conditions that seem to influence the effectiveness of business coaching. Within both CBI and PUM (the organisations that provided the support), the findings helped to inform discussions about targeting future support to specific types of firms and regions

(see van Rijn et al., 2022). The comparative use of the contribution scores by those two support organisations also triggered discussions about the complementarity and synergy of the two business coaching modalities.

We acknowledge some limitations to how the method was used. Aside from the accountability aim, verifying the contributory role of business coaching to improve business performance, the ambition of the PRIME programme was to support adaptive management by implementing staff at CBI and PUM, based on regular (real-time) insights. As discussed in more detail in van Rijn et al. (2022), the level of generalisation of the areas of business management for which the contribution scores were computed limited the learning potential of the research for CBI and PUM implementing staff. More sector-specific versions of the survey, including outcome areas and indicators specific to a certain sector, could have increased the relevance of the data for the staff and experts involved in providing the business coaching support.

Conclusion

In programmes where implementation quality and adaption are key parts of an intervention strategy and where outcomes result from changes in complex systems, there will never be an incontestable answer to the question of how important a specific type of support is or what it contributes to the change process (Mayne, 2019a). The support is, at most, an INUS condition – ‘an insufficient but non-redundant part of a condition which is itself unnecessary but sufficient for the result’ (Mackie, 1974: 62). An evaluation can, however, increase the probability that the support is a non-redundant part of a causal configuration. We argue that contribution scores can help evaluate the impact of private sector development support and permit comparisons of effectiveness between different support modalities on relevant business performance outcomes, and within real-world constraints, and are useful for reflection and learning around the question of What works for whom and under what conditions? (Ton and Vellema, 2022). Contribution scores and perception questions are increasingly used in professional practice to assess the impact of development support. We hope that this paper, with the additional design elements that allow a reflection on response reliability, helps to bolster the method and increase its reliability and acceptance in impact evaluations.

Acknowledgements

The authors acknowledge the inputs and comments of Alex Meerkamp, Thijs van Praag, Jan-Willem Oosterbroek, Rozemarijn Vermeulen, Liesbeth Hofs, Dick de Man, Max Timmerman, and the anonymous reviewers of early versions of this paper. We also want to thank Howard White, Rued Ruben, Robert-Jan Scheer, Yvonne Prince, Cesar Freud, and Jos Walenkamp, who acted as the advisory board for this study. The research was supported by two co-researchers at the Erasmus School of Management, Karen Maas and Job Harms.

Declaration of conflicting interests

Co-funding was provided by the two evaluands, the Centre for the Promotion of Imports from developing countries (CBI) and PUM Netherlands Senior Experts (PUM).

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: from their respective institutions.

ORCID iD

Giel Ton  <https://orcid.org/0000-0003-1730-9517>

Supplementary material

The anonymised data set and statistical analyses used in this paper are available on Figshare 10.6084/m9.figshare.19285433.

References

- Aston T, Roche C, Schaaf M, et al. (2021) Monitoring and evaluation for thinking and working politically. *Evaluation* 28: 36–57.
- Athanasopoulou A and Dopson S (2018) A systematic review of executive coaching outcomes: Is it the journey or the destination that matters the most? *The Leadership Quarterly* 29: 70–88.
- Blackman A, Moscardo G and Gray DE (2016) Challenges for the theory and practice of business coaching: A systematic review of empirical evidence. *Human Resource Development Review* 15: 459–86.
- Crompton BM, Smyrnios KX and Bi R (2012) Measuring the influence of business coaching on fast-growth firms. *Small Enterprise Research* 19(1): 16–31.
- de Mel S, McKenzie D and Woodruff C (2012) *Business Training and Female Enterprise Start-up, Growth, and Dynamics: Experimental Evidence from Sri Lanka*. Washington, DC: World Bank.
- De Meuse KP, Dai G and Lee RJ (2009) Evaluating the effectiveness of executive coaching: Beyond ROI? *Coaching: An International Journal of Theory, Research and Practice* 2(2): 117–34.
- Directorate-General for International Cooperation (DGIS) (2011) *Protocol Resultaatsbereiking En Evalueerbaarheid in PSD*. The Hague: DGIS.
- Directorate-General for International Cooperation–Netherlands Enterprise Agency (DGIS-RVO) (2017) *5 Methodological Notes: Instructions for Calculation, Validation and Reporting of Performance Indicators*. The Hague: DGIS.
- Donor Committee for Enterprise Development (DCED) (2017) The DCED Standard for measuring achievements in private sector development: Control points and compliance criteria. *DCED*. Available at: https://www.enterprise-development.org/wp-content/uploads/DCED_Standard_versionVII_Apr15_bluecover.pdf
- Drexler A, Fischer G and Schoar A (2014) Keeping it simple: Financial literacy and rules of thumb. *American Economic Journal: Applied Economics* 6(2): 1–31.
- Field E, Jayachandran S and Pande R (2010) Do traditional institutions constrain female entrepreneurship? A field experiment on business training in India. *American Economic Review* 100(2): 125–9.
- Funder DC and Ozer DJ (2019) Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science* 2: 156–68.
- Funnell SC and Rogers PJ (2011) *Purposeful Program Theory: Effective Use of Theories of Change and Logic Models*. Hoboken, NJ: John Wiley & Sons.
- Grant AM (2012) ROI is a poor measure of coaching success: Towards a more holistic approach using a well-being and engagement framework. *Coaching: An International Journal of Theory, Research and Practice* 5(2): 74–85.
- Harms J, Ton G and Maas K (2014) *Overview of the literature on the impact of advisory services and export promotion on the performance of small and medium enterprises*. PRIME research brief 2. The Hague: PRIME.
- Humphrey J and Navas-Alemán L (2009) *Multinational value chains, small and medium enterprises, and 'pro-poor' policies: A review of donor practice*. IDS research report 63. Brighton: Institute of Development Studies.
- Lawrence P and Whyte A (2014) Return on investment in executive coaching: A practical model for measuring ROI in organisations. *Coaching: An International Journal of Theory, Research and Practice* 7(1): 4–17.

- Lemire S, Whynot J and Montague S (2019) How we model matters: A manifesto for the next generation of program theorising. *The Canadian Journal of Program Evaluation* 33(3): 53070.
- Mackie JL (1974) *The Cement of the Universe: A Study of Causation*. Oxford: Oxford University Press.
- Mayne J (2001) Addressing attribution through contribution analysis: Using performance measures sensibly. *The Canadian Journal of Program Evaluation* 16: 1–24.
- Mayne J (2011) Contribution analysis: Addressing cause and effect. In: Forss K, Marra M and Schwartz R (eds) *Evaluating the Complex: Attribution, Contribution, and Beyond*. Piscataway, NJ: Transaction Publishers, 53–96.
- Mayne J (2019a) *Assessing the relative importance of causal factors*. CDI practice paper 21, 14 August. Brighton: Institute of Development Studies, Centre for Development Impact.
- Mayne J (2019b) Revisiting contribution analysis. *The Canadian Journal of Program Evaluation* 34(2): 68004.
- Organisation for Economic Co-operation and Development (OECD) (2016) *Private Sector Engagement for Sustainable Development*. Paris: OECD Publishing.
- Pandolfi C (2020) Active ingredients in executive coaching: A systematic literature review. *International Coaching Psychology Review* 15: 6–30.
- Pawson R (2013) *The Science of Evaluation: A Realist Manifesto*. Thousand Oaks, CA: SAGE Publishing.
- Robson PJ and Bennett RJ (2000) The use and impact of business advice by SMEs in Britain: An empirical assessment using logit and ordered logit models. *Applied Economics* 32(13): 1675–88.
- Stern E, Stame N, Mayne J, et al. (2012) *Broadening the range of designs and methods for impact evaluations: Report of a study commissioned by the Department for International Development (DFID)*. Working paper 38. London: DFID.
- Ting S and Hart EW (2004) Formal coaching. In: McCauley CD and Van Velsor E (eds) *The Center for Creative Leadership Handbook of Leadership Development*, 2nd edn. San Francisco, CA: Jossey-Bass, 116–50.
- Ton G (2012) The mixing of methods: A three-step process for improving rigour in impact evaluations. *Evaluation* 18(1): 5–25.
- Ton G and Vellema S (2022) Contribution, causality, context and contingency when evaluating inclusive business programmes. *IDS Bulletin* 53: 1–20.
- van der Windt N, Otgaar A, Heydenreich M, 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, Pamuk H, Dengerink JD, et al. (2022) The search for real-time impact monitoring for private sector support programmes. *IDS Bulletin* 53: 87–102.
- van Rijn F, Ton G, Maas K, et al. (2018a) *Verification of CBI's Intervention Logic: Insights from the PRIME Toolbox*. The Hague: Wageningen Economic Research.
- van Rijn F, Ton G, Maas K, et al. (2018b) *Verification of PUM'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.

Giel Ton works at the Institute of Development Studies. He specializes in contribution analysis for impact evaluation of interventions that support coordination and collective action in agricultural value chains.

Fédes van Rijn works at Wageningen University & Research. She evaluates the impact of agricultural research and innovation, private sector development and sustainability in tropical commodity chains.

Haki Pamuk works at Wageningen University & Research. His research concerns agriculture and food systems innovations and how small and medium-sized enterprises finance those innovations.