Number 17 March 2016

Centre for Development Impact
PRACTICE PAPER
Innovation and learning in impact evaluation

Building Evaluability Assessments into Institutional Monitoring and Evaluation (M&E) Frameworks[1]

Abstract This CDI Practice Paper by Richard Longhurst, Peter Wichmand and Burt Perrin[2] discusses how evaluability assessments (EAs) can support the choice of evaluation approaches for determining impact, drawing on recent experiences of the International Programme on the Elimination of Child Labour of the International Labour Office (ILO-IPEC). These experiences focused on developing a comprehensive monitoring and evaluation (M&E) strategy such that some elements of an EA were built into the system and could be deployed at most points in the programming cycle, in particular to address which questions are important for the evaluation. When used in conjunction with other criteria, this allows for a more informed choice of the evaluation method and related impact.

1 Evaluability assessments: introduction

The importance of evaluability assessment (EA)[3] as a decision-making tool is to assess how to improve design and implementation of an intervention and help decide the nature and timing of evaluation activities. Evaluability has been described as: 'the extent to which an activity or project can be evaluated in a reliable and credible fashion' (OECD-DAC 2010: 21).[4] An EA requires resources and so the benefits must outweigh the cost of carrying it out. Most guidance to date has described EA as a free-standing exercise implemented at the design stage, although recent work has indicated its value across the programming cycle (Davies and Payne 2015; Peersman, Guijt and Pasanen 2015). The key issue is the value and cost of the information it generates for decision-making (whether directly or through an enhanced monitoring and evaluation (M&E) system). This should be considered alongside other factors, including purpose of the evaluation, the existing knowledge base of the implementing organisation, and the level of credibility of that information.

EAs aim to guide the planning, design, implementation and communication of evaluation activities. They should inform judgement about whether an evaluation will provide timely and relevant information for decision-makers; they should help identify what aspects of a programme may be evaluable and when, and indicate the appropriate evaluation approaches or methods to use. It is normally a qualitative investigation that is mostly used before evaluation. These can be cost effective by getting everybody 'on the same page' through the theory of change (ToC) and more 'evaluation aware' during implementation so that the future evaluation is more effective.

EAs have been free-standing; they were used from the late 1970s (Wholey 1979, 1987) but over the past decade have resurged (Davies 2013; Davies and Payne 2015; Trevisan and Walser 2015). Early 1980s literature focused on social welfare programmes in the United States (Schmidt, Scanlon and Bell 1979; IDRC 1996) using several approaches (Trevisan and Huang 2003). Evaluability assessments were used by international finance institutions, often based on assessing value added (Leonard and Eulenburg 2012; IADB 2000) and the main evaluation body within the International Labour Office (ILO), EVAL, has generated numerous EAs (ILO Evaluation Unit 2013). A comprehensive overview of their current and potential use (Davies 2013) recommended their wider application and focused on practical guidance, but raised the possibility of additional costs and procedures, with the danger of them becoming over-extended into the design phase.

Two circumstances were identified that highlighted where the additional costs of an EA might not be justifiable (Davies and Payne 2015): first, when a project design is clearly unevaluable (e.g. the ToC and M&E set-up were unusable). In these circumstances, it is best to address the ToC and M&E framework directly rather than commission a free-standing EA. The second circumstance relates to questionable cost effectiveness for small projects: an EA needs to be cut down to make a cost-effective difference to a potential evaluation.[5] Davies (2013) and Peersman *et al.* (2015) suggest three stages when they could be used – during approval of the intervention design, at approval of the M&E design, and before evaluation – allowing a decision on the purpose, nature and timing of the planned evaluation.

Other options include expanding the process of quality assurance, as used in the Inter-American Development Bank (IADB) to include evaluability and independent procedural audits to examine the functioning of monitoring systems (Davies and Payne 2015). The UK Department for International Development (DFID) and other large institutions have used the procedure of 'approach papers' (starting points for assessment involving literature reviews), which precede the commissioning of evaluations, especially those that are large scale and involve many partners. For small undertakings, they may be appropriate for the initial phase – in essence, an EA.

Traditional evaluation planning often includes EA elements through scoping work and consultation processes with stakeholders to identify methodologies in developing the terms of reference for the evaluation itself. An EA can be included in situation analyses, needs assessment, and 'formative' evaluations in preparation for subsequent evaluations if a second project phase is being considered. Here, there is less scope to use the findings of an EA to ensure the best possible evaluation, but the basis for the choice of evaluation methods could be clarified. Given these viewpoints on the integration of EAs into existing practices, it seems that integrating the procedure in the ongoing M&E process might be a way forward. This paper provides evidence on how it has been used in this way.

2 Evaluation of child labour interventions: key characteristics, M&E approaches and evaluation methods

Factors responsible for child labour are complex, multifaceted and interlinked. Most interventions by the International Programme on the Elimination of Child Labour of the International Labour Office (ILO-IPEC) address two areas: the enabling environment at national and sub-national levels, and targeted interventions (direct actions) with children, families and communities. Efforts to reduce child labour (especially the worst forms) need to address its multiple causes, leading to a range of interventions, including: (1) those stimulated or carried out by ILO-IPEC specifically to address child labour, such as changes in policy and legislation, educational interventions and targeted action with children and families; and (2) other interventions that can influence child labour, e.g. policies and programmes related to education, social protection, livelihoods, employment and labour practice, including occupational health and standards. Contextual factors (such as socioeconomic situation, vulnerability and economic development) are often beyond the control of policymakers and implementers. With these complex interventions operating at many different levels, an evaluation needs to be rooted in a well-crafted ToC operating at many levels and recognising context (see Perrin and Wichmand 2011 for further elaboration).

The current paper uses the practice of ILO-IPEC work on evaluation over the past five years. However, there is a significant track record before that. Over the past 15 years, ILO-IPEC generated considerable experience in the use of different evaluation and research tools to address the complexity of child labour interventions. This has involved more than 250 evaluations at project, programme, thematic and strategic levels, and at mid and full terms. It has also involved using evaluations of first phase interventions to feed into a second phase. Numerous tools and approaches have been tested and refined, (ILO-IPEC 2011a, 2011b) as well as impact evaluations of specific interventions. These tools have generated experience of how and when they could be used, contributing to a strong knowledge base within the organisation.

As this evaluation activity has been coupled with a research programme within ILO-IPEC, there have been very active discussions around questions such as 'what do we need to know?', 'what can we learn?' and 'which tools/techniques do we use to find out what we need to know?' This background conversation lays the basis for subsequent discussion in this paper on 'choice of method' and how the decision-making process can be improved.

Over the past five years, with the increased opportunities for agencies to implement impact evaluation because of stronger interest in the issue on the part of donors, a stronger set of M&E approaches has been required to ensure that impact evaluations were used appropriately and effectively. This includes how they could be implemented using a wider range of evaluative analysis. From this has sprung a lively debate as to the pros and cons of different approaches. It was proposed that impact evaluations should be undertaken to demonstrate impact of a particular type of intervention (used here as a catch-all term for policies, programmes, projects and project components) for the purpose of replication and scaling-up (and as part of the evidence base), rather than to evaluate a project as an institutional modality. Then, definitions of 'impact', the credibility of how it was assessed, and which evaluation techniques were best to use, became important topics for discussion between stakeholders, in particular donor(s) and agency (or agencies).

As a result, a Comprehensive Monitoring and Evaluation Strategy (CMES) (ILO-IPEC 2011b) was developed as an approach to M&E. This was described as 'comprehensive' (or 'integrated') because it attempts to establish and integrate all the elements required to provide evidence of progress (Perrin 2012; Newcomer and Brass 2015). There is greater focus on outcomes and impacts rather than inputs and outputs, including not only on 'what happened' (results) but also addressing 'how' (the implementation process, requirements and time frame) and 'why' changes occurred (with reference to the ToC). The CMES tried to assess the influence of context- and intervention-related factors, and to measure changes outside implementation of interventions – indirect as well as direct interventions, and those implemented by IPEC and by other organisations.

The CMES is seen as a 'strategy' in that it links the elements of the existing M&E system into an integrated and expanded structure, being a much stronger M&E system than is normally found in development projects. But it has the usual steps of an M&E process. Basic elements start with the ToC, design and planning of baseline, institutional assessment and the monitoring of performance. There are two major reasons why a strong M&E (i.e. CMES) is required: first, the focus on outcomes is needed to assess the contribution of individual programmes and projects to the broader national strategy towards eliminating the worst forms of child labour and hence a greater understanding of context. Second, there is a stronger focus on contributing to a knowledge base for stakeholders. A much stronger link between M&E (often weak in many instances) is required.

The CMES approach has been applied to a broad range of projects (21 in all). These range from country-specific projects involving direct action and targeted interventions as well as work on the enabling environment and at policy level, to sub-regional and global research, policy and capacity-building projects. These have included non-ILO projects where ILO-IPEC has served as an external advisor, and projects without ILO support or involvement. The CMES is still in its piloting stage though; guidelines are due to be developed based on the experiences of these projects.

Elements needed for strong M&E form part of the CMES: a well-articulated ToC and monitoring that focuses on linkages and context; an impact or outcome measurement framework (outcomes and indicators); data collection arrangements, especially baseline, end-line or specific data collection during the project; and a sound evaluation system based on the identified methods and outcome-related questions. There is space for keeping stakeholders involved with the ToC and the ongoing M&E processes, through a management and resource plan that updates the CMES based on review meetings.

These are also the elements that provide the information needed to assess the evaluability of the intervention (or parts of it), which can or cannot be evaluated at a given point. The elements are broadly analogous to the verifiable checklists used in free-standing formal EAs and also represent the elements that are needed to inform decisions on what approaches and methods can be used. The CMES process for evaluability provides a well-documented and justified basis, but choices are inevitably muddied by the factors described in Section 3. The EA will be considered in conjunction with these factors, which are general criteria that often arise in decision-making.

The depth of the CMES allows it to carry out an EA at almost any point where it is required. The EA goes beyond addressing the question: is the programme evaluable? That is already agreed. Rather, it assesses (among other things) the strength of the data, the degree of stakeholder support for the ToC in terms of achieving outcomes, required credibility, timing of the evaluation and potential for use. The EA is carried out by the M&E project staff, and involves stakeholders, with support from external M&E advisors, either from the implementing agency or elsewhere. So far, EAs have been carried out to resolve key evaluation questions: extent, choice of methods and use of data.

The M&E process, with key stakeholders (including donors) fully committed, also includes a built-in EA at the various points that the CMES is reviewed. This starts from evaluation set-up and process, including the extent to which impact can be assessed and how, to emerging areas within the project suitable for impact (e.g. areas for learning), to validation of a proposed mid-line, then to assessment of the continued evaluability at the final evaluation. At each point the CMES and its elements will be adjusted if required, including enhancing evaluability. The examples in Section 4 illustrate EA as a built-in process with examples at various points.

Therefore, with these opportunities for assessing evaluability, the CMES does not necessarily include a separate 'EA' step, with a formal process or even report. It is included in the documentation of the review and updating of the CMES. It can replace formal EA exercises such as in the ILO projects, where the requirement for those with a budget of more than US$5m to have an EA is met with the CMES and its built-in review.

However, use of the CMES requires commitment in terms of resources and effort that usually only exists in large projects. In principle, there is no reason why the CMES cannot be used for any level of project – ILO has used it with projects ranging from US$3m to US$15m. The experience with the CMES of using built-in EA has been generated as a result of the internal commitment for M&E to play a major role, especially in circumstances of a pilot or to generate a key learning point. With this, built-in 'assessment of evaluability' rather than a free-standing assessment exercise should always be possible. The next section reviews other criteria that will influence the choice of evaluation methods, which, to some extent, the EA will have taken on board.

3 Criteria that influence the choice of evaluation methods and the role of EAs

EAs can contribute to building an evaluation framework. This has several steps, guiding a decision on what methods could be used. However, there are other criteria for the evaluation manager and commissioner to consider, and these are described below. Within the practical decision-making structure of an evaluation department, trade-offs between these criteria are made subjectively, but the EA will sharpen judgement. More practice with the CMES is needed to draw firm conclusions, but in terms of its role as another element of this decision-making process, it has certain strengths – notably its flexibility, its familiarity with stakeholders, and its quick response to contingencies (such as delays in implementation, or sharp changes in funding).

Purpose, overall framing questions and utility

The position of the intervention in the agency's planning structure and how the evaluation has been initiated are important. There are related issues around the audience, commissioner (e.g. governing body) and roles for stakeholders: whether the evaluation is being carried out largely for accountability or learning purposes, the degree of independence required, how useful it can be, and how far ownership is required within the evaluation process. The CMES should have generated some ownership, which will influence how far the evaluation is participatory. Taking the purpose and evaluation questions on board, the EA will help by indicating which approaches will work.

Level of credibility required of the evaluation results

The level of credibility required in terms of impact (what decision will be made based on the evaluation) is drawn from an EA. If what is needed is a yes/no, 'it works/does not work' decision, a technique providing attribution between causes will provide strong evidence for this. This means an impact evaluation in the form of a randomised controlled trial (RCT) or similar technique, although complementary (mixed) methods will be needed in order to explore the reasons for the findings. If the credibility is based on the extent to which the intervention contributed to the results, then contribution analysis may be used (Mayne 2011). If an independent judgement is needed, then a technique that is less resource-intensive than impact evaluation can be used (by reviewing outputs and outcomes), such as 'plausible association', based on expert judgement, stakeholder interviews, rapid field visits, and an end-of-assignment workshop.

Choice of evaluation method relating to nature of credibility would lead to a standard categorisation of evaluations: what is most often needed is a mix of quantitative and qualitative techniques. Space does not permit a full listing, but the choices include experimental design, qualitative design, theory-based realistic evaluation, participatory evaluation, longitudinal tracking or outcome-focused summative assessment (for more details see Perrin and Wichmand 2011). Different techniques have their own strengths and weaknesses, and often more than one method is needed. An EA will provide, for example, supportive evidence as to whether an experimental design will work, whether stakeholders and decision-makers demand a participatory evaluation, or whether the ToC is rich enough to support a realistic evaluation. Section 4 describes some cases where a CMES helped determine the required levels of credibility.

The level of complexity of the intervention

There is a great deal of literature on complexity, and evaluators are only just coming to terms with what this involves for methodology (Forss, Marra and Schwartz 2011; Befani, Ramalingam and Stern 2015; Bamberger, Vaessen and Raimondo 2016). EAs will illuminate the situation and sharpen the judgement of evaluators. Early experience suggests that the 'running tap' nature of CMES can be a better means of taking complexity on board than a free-standing EA. The key question that managers and evaluators need to address is: what is the level of complexity/reductionism at which an intervention is implemented and an evaluation can be carried out? With a degree of reductionism, do the findings of the evaluation relate sufficiently well to context? Do the findings of the evaluation then provide a sufficient basis for making a decision? If complexity is addressed in design through multiple intervention components, where some involve n=1 (addressed to single actions by governments) and others involve n=thousands (addressed to interventions for children, for example), different evaluation methods can handle this. An EA can help the evaluator decide on the degree of compromise on complexity.

The state of existing knowledge within the commissioning agency

The nature of existing knowledge within the agency contributes to choice of methods. There is no need to repeat evaluations if they do not add to the agency's ability to take decisions. Factors to consider might be: a recent evaluation on a similar topic from which conclusions can be drawn; information banks outside the agency (e.g. systematic reviews, research studies); the need for external validity; the degree of involvement of stakeholders in information-gathering at the design stage; and/or whether worthwhile decisions were made in the past on the basis of existing information that are good enough for sound design. Here, CMES has proved useful to IPEC, showing how it can add to cumulative building of knowledge.

The resources and capacity available for the evaluation

Resources and capacity are often the most important practical criteria influencing the recommendations for an evaluation design; much is determined by the budget line, the managerial capacity of the evaluation team, and willingness to take risks on unfamiliar methods and timelines involved.

Governance arrangements for commissioning the evaluation

The governance structure of the commissioning agency is a key issue as it can inhibit risk-taking for innovative evaluation methods (which also have implications for timelines). If the governing body is conservative, it will steer away from supporting controversial evaluation approaches.

It is quite difficult to say how important each of these factors will be; this depends on each experience, but the EA provides the most objective review of the evidence. There is a political economy of evaluations in agencies such that the most rational choices are not always made. Conflicts over what to do are often resolved by decisions taken by line managers.

4 Operational experiences of using EAs

The experience of ILO-IPEC in carrying out EAs is based on the overall strategic role that M&E plays in the global programme, and the support of technical expertise. The following experiences focus on some choices ILO-IPEC made in approaching impact evaluations. It refers to the CMES elements that were considered most important, and to other factors of EA that might have been addressed.

Assessment of evaluability as part of ongoing operational and strategic planning

Planning and strategy workshops for stakeholders (called Strategic Planning Impact Framework – SPIF) began in 2003 and represent the bedrock of the CMES, allowing an ongoing dialogue to address the ToC. It is a strategic planning process to develop the ToC for interventions, emphasising context and contribution **beyond** the programme. It identifies outcomes at various levels from national, to area of impact (e.g. at sector or defined geographic levels) and programme framework (outcomes and outputs). For each outcome, the indicators and M&E process associated with the interventions leading to that outcome are identified, and so a complete M&E framework emerges.

These stakeholder workshops, combined with studies and documentation, are part of the national planning process and of project design. As the project is approved and the M&E framework is developed, the operational planning process includes further workshops to validate and adjust the design and the M&E system, including logframes, project results framework and monitoring, and the scope and possibility for evaluation. If the project is properly implemented, and design has used a proper stakeholder engagement, then CMES will identify the basis for determining the degree of evaluability of the intervention.

The workshops are carried out at regular intervals, considering the ToC, outcomes, results and contextual factors. They constantly look at the nature of the intervention and the level of complexity and the implications for the evaluation. Strong facilitation and technical M&E skills are needed to effectively lead SPIF workshops. They bring some well-documented evidence to the purpose of the evaluation, evaluation questions and utility for the choice of evaluation approaches and the intended levels of credibility.

Assessment of evaluability built into the M&E process

The following country examples show how key decisions on evaluation – including choice of approaches and methods and data – were based on analyses as part of the CMES.

Ghana: choice of data and level of credibility

In 2010, ILO-IPEC rolled out a four-year project in Ghana, the Cocoa Communities Project. This aimed to eliminate child labour in cocoa-growing communities by strengthening community action and social surveillance, enhancing stakeholders' efforts, and by improving household livelihoods and children's access to education. The CMES was piloted here, and generated an EA. It worked as a learning tool, leading to information-sharing, media strategy and consultations. This resolved competing ideas for the use of the baseline and also identified the combination of evaluation methods.

As part of the CMES, stakeholders generated a shared ToC; the CMES was well integrated with the intervention and had strong stakeholder involvement. Data on other actors were included with the context monitoring, which made staff aware of how to avoid duplication and helped implementing agencies to identify partners to support project effectiveness. As noted in Section 2, the CMES is a more extensive M&E system than is normally used, so initially, it appeared complex. But learning-by-doing changed attitudes, with the validation of indicators identified during the planning phase taking a year. One of the initial lessons learned from using the CMES was the limited opportunities for obtaining full benefits of information-sharing unless this was explicitly included in the intervention design and M&E. Information-sharing needs to be facilitated throughout, with regular interactions with stakeholders, in order to support sustainability.

The CMES identified the evaluation methods as a combination of: required mid-term evaluation; an impact evaluation with counterfactual; and a final evaluation with detailed sub-studies (known in IPEC as Expanded Final Evaluation – EFE) on the livelihood component. These were the methods proposed for the full package of integrated area approach intervention that was being developed and pilot tested. As the project was implemented and the CMES provided information on progress, it suggested that a project implementation review would be enough for the mid-term evaluation as the impact evaluation and EFE would provide fuller testing of the ToC. The evaluations for Ghana therefore consisted of a project implementation review, impact evaluation with implemented baseline and EFE.

The impact evaluation was seen as an element of the EFE, with a challenge to integrate the impact evaluation from the start in the M&E strategy. In Ghana, the impact evaluation approach was experimental and assessed the whole intervention package. These approaches were decided based on key stakeholder demands, with national stakeholders needing evidence about what works in child labour (how and why) to understand its complexity.

Stakeholders needed to know how to replicate and scale-up, and which baseline to use. The government was already collecting baseline data and was insisting that these data be used (for consistency and replication) rather than the baseline data collected as part of the CMES. However, the EA stemming from the CMES suggested that to enable methodological consistency, timing, coverage and quality of data, it was more appropriate to use the baseline data collected specifically for the project. This decision was helped by earlier agreements generated by the CMES on the ToC and the framework, which had included identifying indicators and their collection at outcome level for national and local government agencies. The validation of the CMES during implementation generated the basis to understand the 'how'.

El Salvador: choice of approach to impact studies – overall framing questions

In El Salvador, the project focused on improving livelihoods, providing direct support to schools and sensitising them on the dangers of child labour. There was expected to be closer integration of child labour concerns with government anti-poverty and social inclusion policies and programmes. The project worked with an integrated area-based approach and a simplified CMES at local level. The challenge has been to integrate the impact evaluation from the start with the M&E strategy; the impact evaluation is quasi-experimental and focused on three interventions.

The EA that formed part of the CMES provided justification for key decisions at the design stage about which interventions the evaluation should cover, and on choice of approach to impact studies and overall framing questions and utility. A number of potential sub-studies for the EFE were identified, using CMES information and ongoing dialogue, including through the mandatory mid-term evaluation. The tool that worked best here was concept notes prepared for discussion by M&E specialists in consultation with stakeholders, leading eventually to the selection of the topics for sub-studies and also highlighting issues of timing and scope. The concept notes provide an example of an EA carried out internally during the implementation process that helped to establish, after stakeholder discussion, the specific evaluation methods – one of which included quantitative repeat baseline study. As a result, the agreed evaluation framework involved an EFE with sub-studies that complemented and fed into it. Project beneficiaries, follow-up/repeat baseline and the impact evaluation were regarded as specific knowledge products that took on the role of (and replaced) the sub-studies components normally part of an EFE.

EA for specific evaluations in a broader M&E framework

Thailand: decision on feasibility of impact evaluation – level of credibility based on sample size

The Thailand project started in 2010 and focused on the shrimp processing sector, involving migrant children, for whom official identification was an issue. The donor was pressing for an RCT, and so the project design included provision for an impact evaluation using experimental or quasi-experimental methods. The impact evaluation was only included at the very end of the design process, in 2009, and therefore was not fully integrated into the original design. A CMES and impact evaluation plan were developed, detailing M&E activities, including the impact evaluation (which was not actually carried out).

The baseline study was redesigned to provide documentation for the incidence of child labour as a requirement by the government to proceed. It could not therefore serve the purpose of the evaluation, so the intention was to carry out an additional impact evaluation-focused baseline that was technically and methodologically possible, based on the information in an EA, generated through the CMES. Analysis of available data indicated that sufficient numbers could not be reached for the sample size to make the evaluation design feasible, given the requirement for an RCT. Other, qualitative approaches using contribution analysis were considered as part of the EFE, but were also not considered feasible.

The assessment addressed whether sufficient numbers of control and intervention groups could be reached by the project to make the analysis possible. The evaluation design was applicable to the proposed intervention; while the intervention related to migrant population and the evaluation design took that into account, it was not carried out because the experimental design (to be funded by the donor) required sample sizes for intervention groups that were not feasible. Some factors had also changed significantly and a constant re-assessment was required to address the nature of the final evaluation. As an integrated process, the CMES was able to address stakeholder needs in a way that a standalone EA could not do.

Peru and the Philippines: identifying and designing most appropriate impact evaluation – state of existing knowledge within the commissioning agency

In 2011, ILO-IPEC served as the external M&E advisor to two US Department of Labor-funded non-ILO implemented projects to design and implement impact evaluations in Peru and the Philippines. These projects needed a strong design and needs assessment, with documentation of proposed interventions, leading to ready-to-start interventions. Timing and delivery requirements were very explicit and demanding, as these were seen by the donor as the key performance criteria

for assessing funding for the implementing agency. This experience provides some lessons on inserting impact evaluations into already designed projects and using EA as the basis for identifying an appropriate evaluation. The EA helped design the most appropriate impact evaluation, strengthened the level of credibility and enhanced the level of knowledge within the commissioning agency.

Conclusions drawn from these two experiences point to the obvious need to design impact evaluations before or while the intervention itself is being designed, ensuring that the selection of treatment and comparison groups match the design and planned roll-out of the programme. It also ensures that political leverage for preserving its validity still exists: when political decisions have been made and publicised, negotiating changes for the sake of the impact evaluation will no longer be possible (or will be highly contentious).[6] But achieving change around the incidence of child labour takes time: in Peru and the Philippines, securing change took at least 12–24 months, also depending on education and agricultural cycles.

In terms of deciding which interventions were to be part of the impact evaluation, the EA considered issues such as: the potential for replication and scaling-up interventions with sufficient beneficiary numbers to sustain transparent and clear rules of allocation; the feasibility of developing a clear ToC for the intervention; and assessing where an observable change is expected within the time frame of the evaluation design. Finally, if possible, an EA should identify interventions where the time frame of the evaluation study can be continued beyond the time frame of the initial design in order to evaluate longer-term impacts. In this case, the impact evaluation proceeded but had to take into account numerous issues, mostly around sample size, which were illuminated by the CMES. The rationale for carrying out impact evaluations was to expand the knowledge base for the donor, sometimes to the detriment of carrying out activities immediately. It requires a commitment that knowledge will mean better decisions and better investments in the fight against child labour.

Although the knowledge generated through an impact evaluation is an output that contributes to the knowledge base, the project's objectives do not usually recognise this as one of its outcomes. Project managers are required to prioritise delivery of the stated project outcomes. The importance given to the implementation of the impact evaluation and its contribution to the knowledge base could be improved by including the evaluation as one of the expected outcomes (as also noted by Shutt and McGee 2013).

5 Conclusions

This paper has described experiences of using evaluability assessments within the complex child labour evaluations of ILO-IPEC's interventions. A comprehensive M&E strategy was developed that allowed EAs as an integral part of the interventions. Many of the impact evaluations that were subject to an EA have only been concluded recently, and so the learning generated requires further experience. Emerging lessons suggest that there are some key elements: a well-tested ToC with stakeholder workshops, strong monitoring of context, and development of outcome measurement frameworks. The CMES provided flexibility in carrying out EAs, especially in response to contingencies and differing views. It also provided integrated evidence that has more ownership by stakeholders, more so than external EAs. Transaction costs are likely to be lower than an external assessment. Finally, we highlight three points.

- ILO-IPEC experience shows that it is worth considering integrating EAs into existing M&E set-ups, in terms of improving the ToC and stakeholder engagement, and being able to deal with contingencies. Integrated EAs may also be more effective in dealing with complexity.
- EAs may allow early planning for evaluations, which can make stakeholders more 'evaluation aware'. Built-in EAs also allow the knowledge base of the agency to be developed more systematically, and enable better learning and information-sharing among stakeholders. Periodic revisiting of the ToC will allow for better implementation.
- There may be cost factors to consider, and these need to be weighed (subjectively) against the benefits of conducting an EA. One of the benefits would be capacity building of M&E staff, national stakeholders and others involved in the CMES.

Notes

1 The views expressed in this paper are personal and do not reflect the view of the organisations with which the authors are affiliated. The authors wish to thank Rick Davies, Kathi Welle and Chris Barnett for their most useful and detailed comments.

2 Respectively, Research Associate, Institute of Development Studies, Brighton; Senior Evaluation Officer, ILO Evaluation Office; formerly Head of Evaluation and Impact Assessment Unit, International Programme on the Elimination of Child Labour, International Labour Organization, Geneva; and Independent Consultant, Vissec, France.

3 Evaluability assessments can be a clearly defined exercise and here we refer to these as 'free-standing'. However, in all of the ILO-IPEC examples, the EAs are built into the ongoing M&E process (with constant assessment of evaluability), which is called a Comprehensive M&E Strategy (CMES).

4 See OECD (2010). A rather more narrow definition has been provided by the Evaluation Cooperation Group of the international financial institutions (and used in an EBRD Working Paper), which is: '*the extent to which the value generated or the expected results of a project are verifiable in a reliable and credible fashion*'. The World Bank uses the following definition: '*A brief preliminary study undertaken to determine whether an evaluation would be useful and feasible... It may also define the purpose of the evaluation and methods for conducting it.*'

5 Where the intervention to be evaluated is a pilot intended for replication and scaling-up, it is very justified to put considerable efforts into the evaluation. Indeed, the evaluation is considered an integral component for the project to deliver and such a pilot could not be justified without considerable evaluation effort.

6 Vermeersch, Rothenbühler and Renee Sturdy (2012).

References

Bamberger, M.; Vaessen, J. and Raimondo, E. (2016) *Dealing with Complexity in Development Evaluation: A Practical Approach*, SAGE Publications

Befani, B.; Ramalingam, B. and Stern, E. (2015) 'Introduction – Towards Systemic Approaches to Evaluation and Impact', *IDS Bulletin* 46.1: 1–6

Davies, R. (2013) *Planning Evaluability Assessments: A Synthesis of the Literature with Recommendations*, DFID Working Paper 40, London: Department for International Development, www.gov.uk/government/uploads/system/uploads/attachment_data/file/248656/wp40-planning-eval-assessments.pdf (accessed 15 January 2016)

Davies, R. and Payne, L. (2015) 'Evaluability Assessments: Reflections on a Review of the Literature', *Evaluation* 21: 216–31

Forss, K.; Marra, M. and Schwartz, R. (eds) (2011) *Evaluating the Complex: Attribution, Contribution and Beyond*, New Jersey: Transaction Publishers

ILO Evaluation Unit (2013) *Annual Evaluation Report 2012–2013*, Geneva: ILO, www.ilo.org/global/docs/WCMS_226358/lang--en/index.htm (accessed 29 January 2016)

ILO Evaluation Unit (2011) *Using the Evaluability Assessment Tool*, Guidance Note 11, www.ilo.org/wcmsp5/groups/public/---ed_mas/---eval/documents/publication/wcms_165984.pdf (accessed 25 January 2016)

ILO-IPEC (2011a) *Child Labour Impact Assessment Toolkit*, International Labour Organization, International Programme on the Elimination of Child Labour (IPEC), Geneva: ILO, www.ilo.org/ipec/Informationresources/WCMS_IPEC_PUB_8990/lang--en/index.htm (accessed 25 January 2016)

ILO-IPEC (2011b) 'Initial Generic Guidance Notes on Comprehensive Monitoring and Evaluation Strategy (CMES)', *Child Labour Impact Assessment Toolkit*, International Labour Organization, International Programme on the Elimination of Child Labour (IPEC), Geneva: ILO

ILO-IPEC (2011c) 'Tracer Study Manual', *Child Labour Impact Assessment Toolkit*, International Labour Organization, International Programme on the Elimination of Child Labour (IPEC), Geneva: ILO, www.ilo.org/ipec/Informationresources/WCMS_IPEC_PUB_8990/lang--en/index.htm (accessed 25 January 2016)

International Development Research Center (IDRC) (1996) *Guidelines for Preparing an Evaluability Assessment*, IDRC, http://web.idrc.ca/en/ev-29113-201-1-DO_TOPIC.html (accessed 15 January 2016)

Inter-American Development Bank (IADB) (2000) *Evaluability Assessment in Project Preparation*, Washington DC: Office of Evaluation and Oversight

Leonard, K. and Eulenburg, A. (2012) *Evaluability – Is It Relevant for EBRD?*, Evaluation Brief, European Bank for Reconstruction and Development, Evaluation Department, www.ebrd.com/downloads/about/evaluation/130214.pdf (accessed 15 January 2016)

Mayne, J. (2011) 'Contribution Analysis: Addressing Cause and Effect', in K. Forss, M. Marra and R. Schwartz (eds), *Evaluating the Complex: Attribution, Contribution and Beyond*, New Jersey: Transaction Publishers

Newcomer, K. and Brass, C.T. (2015) 'Forging a Strategic and Comprehensive Approach to Evaluation within Public and Nonprofit Organizations: Integrating Measurement and Analytics within Evaluation', *American Journal of Evaluation* DOI: 1098214014567144

OECD (2010) *Glossary of Key Terms in Evaluation and Results-Based Management, Evaluation and Aid Effectiveness No 6*, Paris: OECD Publishing

Peersman, G.; Guijt, I. and Pasanen, T. (2015) *Evaluability Assessment for Impact Evaluation: Guidance, Checklists and Decision Support*, Methods Lab, London: Overseas Development Institute, www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/9802.pdf (accessed 15 January 2015)

Perrin, B. (2012) 'How to Take a Strategic Approach to Evaluation That Can Contribute to Performance Improvement and Development Impact', workshop at the AfrEA conference, Accra, January 2012 and at the EES conference, Helsinki, October 2012

Perrin, B. and Wichmand, P. (2011) 'Evaluating Complex Strategic Development Interventions: The Challenge of Child Labour', in K. Forss, M. Marra and R. Schwartz (eds), *Evaluating the Complex: Attribution, Contribution and Beyond*, New Jersey: Transaction Publishers

Schmidt, R.; Scanlon, J. and Bell, J. (1979) *Evaluability Assessment: Making Public Programs Work Better*, Rockville: Project SHARE

Shutt, C. and McGee, R. (2013) *Improving the Evaluability of INGO Empowerment and Accountability Programmes*, CDI Practice Paper 1, Brighton: IDS, www.ids.ac.uk/publication/improving-the-evaluability-of-ingo-empowerment-and-accountability-programmes (accessed 25 January 2016)

Trevisan, M. and Walser, T. (2015) *Evaluability Assessment: Improving Evaluation Quality and Use*, California: SAGE Publications

Trevisan, M. and Huang, Y.M. (2003) 'Evaluability Assessment: A Primer', *Practical Assessment, Research & Evaluation* 8.20: 2–9

Vermeersch, C.; Rothenbühler, E. and Renee Sturdy, J. (2012) *Impact Evaluation Toolkit, Measuring the Impact of Results-Based Financing on Maternal and Child Health*, Version 1.0. Washington DC: World Bank, http://siteresources.worldbank.org/INTIMPEVALTK/Resources/IE_Toolkit_2012.08.21_ENG.pdf (accessed 25 January 2016)

Wholey, J. (1987) 'Evaluability Assessment: Developing Program Theory', *New Directions for Programme Evaluation* 33: 77–92

Wholey, J. (1979) *Evaluation: Promise and Performance*, Washington DC: Urban Institute

Centre for Development Impact (CDI)

The Centre is a collaboration between IDS (www.ids.ac.uk), Itad (www.itad.com) and the University of East Anglia (www.uea.ac.uk).

The Centre aims to contribute to innovation and excellence in the areas of impact assessment, evaluation and learning in development. The Centre's work is presently focused on:

(1) Exploring a broader range of evaluation designs and methods, and approaches to causal inference.
(2) Designing appropriate ways to assess the impact of complex interventions in challenging contexts.
(3) Better understanding the political dynamics and other factors in the evaluation process, including the use of evaluation evidence.

This CDI Practice Paper was written by **Richard Longhurst**, **Peter Wichmand** and **Burt Perrin**.

ISSN: 2053-0536
AG Level 2 Output ID: 619

Institute of Development Studies, Brighton BN1 9RE, UK
T +44 (0) 1273 915637 **F** +44 (0) 1273 621202 **E** ids@ids.ac.uk **W** www.ids.ac.uk