Governing Body

316th Session, Geneva, 1–16 November 2012

GB.316/PFA/7/1

Programme, Financial and Administrative Section
Audit and Oversight Segment

PFA

Date: 5 October 2012
Original: English

SEVENTH ITEM ON THE AGENDA

Annual evaluation report 2011–12

Purpose of the document

The present document provides an overview of the work undertaken by the Evaluation Unit (EVAL) during 2011–12 as measured against its results-based strategy 2011–15 and a segment on the Office's effectiveness in achieving short- and medium-term objectives. The Governing Body is invited to take note of the report, to endorse recommendations to be included in the ILO's rolling plan of action for the implementation of the recommendations and suggestions contained in the annual evaluation report and to confirm priorities for the programme of work (see the draft decision in paragraph 96).

Relevant strategic objective: All.

Policy implications: The implementation of the recommendations of each evaluation may have policy implications.

Legal implications: None.

Financial implications: None.

Follow-up action required: The ILO's Evaluation Unit will incorporate approved recommendations into its rolling action plan for the implementation of the recommendations and suggestions contained in the annual evaluation reports (Appendix I).

Author unit: Evaluation Unit (EVAL).

Related documents: GB.313/PFA/7/1; GB.312/PFA/8; GB.312/PFA/9; GB.310/PFA/4/1(Rev.).

Contents

Appendices

Introduction

1. Each year, EVAL presents an annual report on the progress made in implementing the ILO's evaluation policy and strategy to the Governing Body. Last year's report (2010–11) represented a major change in terms of content and structure, as it had to reflect the new results-based strategy for the evaluation function and respond to the Governing Body's request for more information on the Office's overall effectiveness in achieving short- and medium-term objectives.

2. This year's report (2011–12) continues to use the revised format, reporting on the progress made under the three outcomes identified in the strategy, the obstacles encountered and new knowledge that will be applied to future work and reports. As it is a mid-biennium report, no concrete results on the biennial milestone targets can be provided but an indication of expected results is included. The report also covers actions taken regarding the plan of action for the implementation of the recommendations and suggestions contained in the annual evaluation report 2010–11. Appendix I explains the work done and the next steps being taken by the Office.

3. In a climate of austerity and growing concern about value for money, international agencies and development partners are subject to increasing pressure to show the impact and effectiveness of their interventions. Therefore, developing and demonstrating a stronger evaluation infrastructure and culture within the ILO has been an important part of the Office's approach. As a normative agency with a long tradition of collaboration in setting and monitoring the implementation of international standards at the country level, the ILO has a significant advantage over other United Nations agencies in terms of monitoring and evaluation. The ILO is recognized for its leading role in promoting and achieving the Decent Work Agenda through specialized technical expertise. However, it also has the potential to improve its reputation considerably by expanding evaluation and impact assessment in these areas to benefit its constituents, who look to the ILO for knowledge products and tools.

4. As the report illustrates, the ILO made notable strides in strengthening the evaluation function. These improvements (increased resources, enhanced independence, etc.) have cemented the evaluation infrastructure of the Office *(the hardware)*. This now needs to be matched with the continued fine-tuning of the Office's culture *(the software)* in order to fully appreciate evaluation as a tool for management and for organizational learning at all levels. The use of evaluation results for governance and management purposes has improved but requires further reinforcement. EVAL will continue its efforts to support line managers by providing meaningful follow-up to evaluation recommendations.

5. The current evaluation policy reinforces the credibility of ILO evaluations, enabling them to make a greater contribution to improving the effectiveness of ILO programmes and projects. In stepping up to the challenge of finding ways to provide more credible evidence on the outcomes and cost effectiveness of policies, programmes and interventions, the ILO needs to generate more and better evidence on the impact of its normative and policy work.

Part I. Implementation of the ILO's 2011–15 evaluation strategy

A. Progress made towards achieving key strategy milestones

6. Following the Governing Body's decision to approve the evaluation strategy in March 2011, a report on its implementation is presented as part of the annual evaluation report. The results matrices set for each strategy outcome are provided in Appendix II.

Outcome 1: Improved use of evaluation by management and constituents for governance

7. The ILO evaluation strategy emphasizes the role that evaluation plays in supporting senior management and the Governing Body in exercising accountability to ensure the ILO's relevance, effectiveness and impact. The strategy calls upon the Evaluation Advisory Committee (EAC) to facilitate the strategic use of evaluation. Likewise, the strategy aims to improve the strategic choice of evaluation topics, to validate the quality and credibility of the evaluations, and to take stock of evaluation results in order to gain a better understanding of the ILO's development and operational effectiveness. The progress made within these strategy components is outlined below.

B. Improving the effectiveness of the Evaluation Advisory Committee

> *Biennial milestone 1.1 (2012–13): Four meetings per year; record of recommendations for evaluation programme of work; record of EAC advice on use of specific recommendations.*

8. The Office has committed to strengthening the EAC through regular meetings and a more substantive agenda to catalyse the use of evaluation. Following a good start in early 2011, the EAC has been without a chair since September 2011. Nevertheless, several virtual meetings have taken place in order to advance discussions on key agenda items including the external quality review of high-level evaluations, the choice of high-level evaluation topics, and the progress on the implementation of the 2011 high-level evaluations.

9. Once the new EAC chair is appointed by the Director-General, EVAL will work closely with the chair to clear the backlog of agenda items, which will include a more critical review of the adequacy of the follow-up given to evaluation by ILO line managers.

C. Assessing ILO performance

> *Biennial milestone 1.2 (2012–13): Improved annual evaluation report based on Governing Body feedback; results are fed into the Programme and Budget for 2014–15.*

10. The revised reporting format of 2011 added a new part to the annual evaluation report. This part highlights the results of evaluative studies on the Office's overall effectiveness and achievement of results. This year, Part II of the report continues to look at results-based management in the Office and the overall performance of activities funded by the Regular Budget Supplementary Account (RBSA). It also reviews the effectiveness of

quality appraisal systems introduced by the Office for new Decent Work Country Programmes and technical cooperation, as well as that of performance rating practices in the ILO and how they can be improved. Next year (and for each odd-numbered year of the programme and budget biennium), EVAL will again take stock of the information regarding performance contained in project and programme evaluations in order to provide those planning the ILO's new programme of work with some insight.

D. Follow-up to high-level evaluations

11. The adequacy of the follow-up to the 2011 high-level evaluations is a major area of scrutiny for the EAC. Table 1 summarizes the progress reported and EVAL's determination of the recommendations successfully implemented from the evaluations of the Bahía Decent Work Agenda (Brazil), and the ILO's strategies to address HIV and AIDS and the world of work, and to eliminate discrimination in employment and occupation. Of the 21 recommendations, EVAL considers 11 to have been fully implemented, seven to have been partially implemented and one not to have been accepted. Once a chair has been appointed, the EAC will be requested to confirm EVAL's assessment that, in all cases, satisfactory progress has been made.

Table 1.   Summary of planned follow-up and completion status,
2010 high-level evaluations

Evaluation topic	Total recommendations	Action planned	Completed	Partially completed	No action taken
Bahía Decent Work Agenda (Brazil)	5	4	0	4	1
ILO strategy to address HIV and AIDS and the world of work	10	10	7	3	0
ILO strategy for the elimination of discrimination in employment and occupation	6	6	4	2	0

12. Highlights related to management's use of and follow-up to the evaluations include:

■ Follow-up to the evaluation of the ILO strategy to address HIV and AIDS and the world of work included the submission of an action plan to the Committee on Employment and Social Policy (ESP) in March 2012, which addressed many of the evaluation's recommendations.

■ Management has taken specific actions regarding the six recommendations of the evaluation of the ILO's strategy for the elimination of discrimination in employment and occupation. The findings and recommendations provided a reference framework for the recurrent discussion of the International Labour Conference (ILC) on fundamental principles and rights at work in June 2012, which resulted in an action plan.

■ In 2012, Brazil launched a series of 27 state conferences on job creation and decent work in an effort to promote social dialogue. The state conferences culminated in the formulation of proposals, and in the appointment of delegates for the National Conference on Employment and Decent Work, which was convened and held by the Government of Brazil in Brasilia in May 2012.

13. Summaries of the Office's 2012 high-level evaluations have been submitted to the Governing Body (GB.316/PFA/7/2); they cover the ILO's strategy to support inclusive employment policies (programme and budget outcome 1.1), decent work at the sectoral level (outcome 13) and India's Decent Work Country Programme.

E.   Independent quality review of high-level evaluations

> *Biennial milestone 1.3 (2012–13): Results of internal peer review of high-level evaluations 2012−13 register satisfactory quality.*

14. In the conclusions of the *Independent External Evaluation of the International Labour Office Evaluation Function,* 2010, the consultants recommended that the Office regularly submit high-level evaluations for independent peer review in order to improve their quality and usefulness. The Office adopted this recommendation and included an indicator and biennial milestones on the quality of high-level evaluations in the revised results-based strategy for evaluation.

15. On EVAL's recommendation, the EAC agreed to upgrade the exercise from a peer review to an independent review for the 2010–11 high-level evaluations and to oversee the process to ensure independence.

16. The validity of high-level evaluations is dependent upon sound evaluation design, independent, transparent and professional processes, and upon a report that is both of a high quality and useful. The independent review assessed the quality of each evaluation process, report and the follow-up given thereto.

17. The review rated the quality of reports from medium to high quality, and found all to be fair and reasonable assessments of the programmes being evaluated. It was also noted that the evaluation findings were mostly used within the Office, usually in the form of knowledge gained and applied by the programme evaluated. Feedback also suggested that opportunities existed for the greater involvement of internal and external stakeholders in the process. Overall, all reports were found to be weak in addressing the longer term development impacts of the various programmes. The analysis and reporting of good practices as a separate section was also found to be weak.

18. As shown in figure 1 below, no specific part of the evaluations was consistently of a low quality.

Figure 1.   Average rating of quality by evaluation report sections
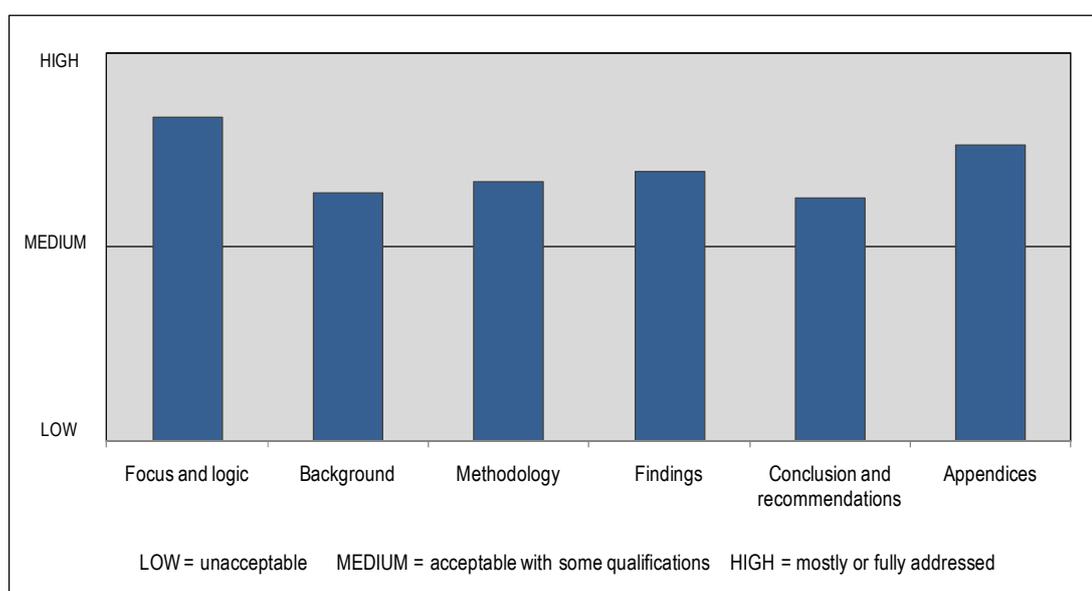
F.   Selecting high-level evaluation topics
for strategic use

19. Each year, the Director of Evaluation seeks input from senior management, members of the EAC and the ILO's constituents on the high-level topics to be evaluated over the next three years. The results of these online consultations determine the rolling workplan for proposed evaluations, which is submitted to the Governing Body each November. This year, the Office and the constituents discussed the possibility of postponing the evaluation of the field structure reform until 2015. Several groups of constituents suggested that the evaluation of the technical cooperation strategy be brought forward to 2013 instead. Table 2 shows the evaluation topics agreed upon by the Governing Body and those to be confirmed.

Table 2.   Confirmed high-level evaluations for 2013,
and those proposed for 2014 and 2015

Year	Evaluation type	Topic of independent evaluation	Timing	Rationale
Decided 2013	Strategy	Workers and enterprises benefit from improved safety and health conditions at work	2013	Evaluation last discussed in 2002. Endorsed as a topic in November 2011
Proposed 2013	Strategy	Sustainable enterprises create productive and decent jobs (to replace field structure review originally planned for 2013)	2013	Evaluation last discussed in 2003 (can follow-up on 2012 GB discussion)
Decided 2013	Decent Work Country Programmes	Comparative evaluations of Decent Work Country Programmes in Arab States	2013	Last discussed in 2008. Endorsed as a topic in November 2011
Proposed 2014	Strategy	Integrated approach to economic and social policies supported by UN and multilateral agencies	2014	Not yet evaluated; proposed by several constituents
Proposed 2014	Strategy	ILO's action on fundamental principles and rights	2014	Pre-selected by ILC for completion by June 2015
Proposed 2014	Decent Work Country Programme	Africa region	2014	Last discussed in 2010; North Africa
Proposed 2015	Institutional capacities	ILO's Technical Cooperation Strategy	2015	Proposed by several constituents and senior ILO management
Proposed 2015	Institutional capacities	Field structure review, including involvement of governments and employers' and workers' organizations in Decent Work Country Programmes (postponed due to planned changes to field structure in 2013)	2015	Governing Body-mandated evaluation to review progress/effectiveness of field structure review – postponement from 2013 to 2015 suggested in light of additional changes to field structure being considered
Proposed 2015	Decent Work Country Programme	Americas	2015	Last discussed in 2009; Central America

Outcome 2: Harmonized Office-wide evaluation practice to support transparency and accountability

Harmonizing and standardizing types of evaluations
and associated roles and responsibilities to improve
value and efficiency

20. **Codify and upgrade procedures and guidelines**: Completing a comprehensive resource kit on the ILO's evaluation policy that included related procedures, guidelines and guidance notes based on good practice was an important goal for EVAL in its efforts to strengthen consistency in evaluation practices throughout the Organization. EVAL has finalized its new product called "*i-eval*: Resource kit", which serves as a gateway to ILO expertise and knowledge on evaluation. The resource kit has proven highly popular, with over 1,000 copies printed and regular downloads and "clicks" to access it via the ILO Intranet. As set out in the evaluation strategy, EVAL has also completed user-friendly policy guidelines and more detailed guidance on two new evaluation areas, namely impact evaluation and joint evaluation.

21. **Impact evaluation**: The ILO uses impact evaluation to support governments and employers' and workers' organizations in determining the effects of various decent work policies within specific countries; to provide valuable feedback to the ILO on the extent to which its policy support, programmatic approaches and technical tools are appropriately designed and delivered; and, at the project or intervention level, to test the credibility and validity of new concepts and techniques that still need to be proven.

22. Impact evaluations do not directly support the Office's accountability framework and are not funded by the current ILO regular budget. As a result, there is no required reporting to the Governing Body on impact evaluations. For quality control, EVAL will carry out a quality review at the design and completion stages if impact evaluations are to be discussed by the Governing Body. Impact evaluations should be commissioned strategically in order to assess those policies and programmes that are unproven and that are either widely used or are new and promising.

23. **Joint evaluation**: The number of joint evaluations involving the ILO has dramatically increased in recent years (see figure 2 on types of evaluations below). EVAL estimates that the ILO is actively involved in 30 joint evaluations each year. Nearly all of these are linked to the United Nations Development Assistance Framework (UNDAF) and/or joint programmes with the United Nations or partners of international financial institutions. According to the Development Assistance Committee (DAC) of the Organisation of Economic Co-operation and Development (OECD), the definition of a joint evaluation is a development evaluation that is conducted collaboratively or jointly by more than one agency.

24. The ILO will proactively participate in the joint evaluations, not only as a lead agency, but also as a partner agency. It is a cost-efficient opportunity to obtain findings that are relevant to the ILO as an agency within a joint agency initiative. ILO staff are strongly encouraged to participate in evaluation management if the ILO component of the project surpasses the threshold of US$500,000. At US$1 million their participation is compulsory. All joint evaluation reports that cover the ILO's work, even partially, should become part of its evaluation knowledge base.

G. Upgrading and expanding the use of decentralized evaluations

25. Independent project evaluations increased from 71 in 2010 to 96 in 2011. This 26 per cent increase was mainly due to the ILO's participation in joint programmes. Of the 96 independent evaluation reports received, ILO staff managed 43 while the remaining 53 were managed externally (see figure 2). Of the externally managed evaluations, the United Nations joint programmes managed 47 and the European Commission (EC) or a consulting company managed the remaining six. The increase in joint evaluations can be attributed to the fact that mid-term evaluation reports became due for the many projects under the Millennium Development Goals Achievement Fund, to which the ILO contributes. EVAL is collecting all of these evaluations, regardless of budget threshold.

Figure 2.   Distribution by type of independent evaluation, 2007–11
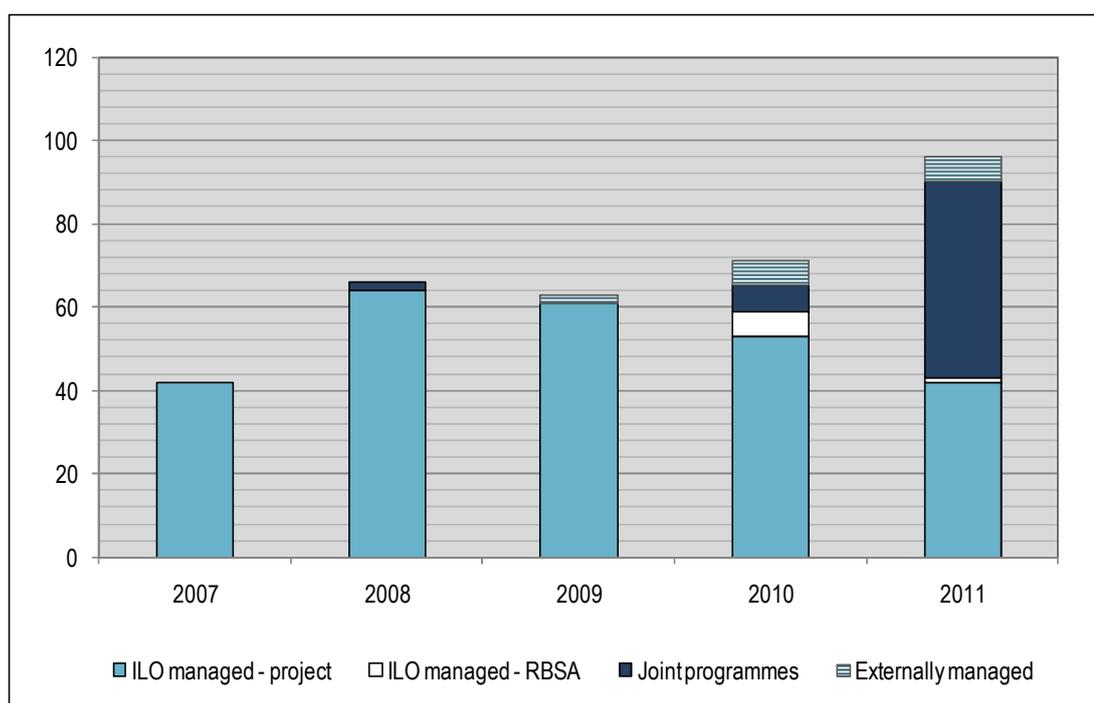
26. Of this year's independent evaluations, 40 were final and 56 were mid-term evaluations. A full list of these evaluations is available in the long version of this report, which can be found on the ILO's website (www.ilo.org/eval).

27. Of the 43 independent evaluation reports managed by the ILO in 2011, 32 were included in the management follow-up exercise. [1] The remaining evaluations did not require a follow-up exercise (11). The 32 evaluations from 2011 that were subject to the follow-up exercise are summarized in table 3 below, which indicates the response rate by line managers and the status of follow-up by region. As can be seen, four plans for follow-up were not received by the time this report was written.

---

[1] This year, four evaluations managed by the ILO International Programme on the Elimination of Child Labour (IPEC) have been included in the follow-up reporting. Last year's report erroneously excluded evaluations managed by IPEC in the follow-up reporting.

Table 3.   Management response to evaluations completed in 2011

Region/sector	Management responses (32 reports)		Total recommendations received	Completed	Partially completed	Outstanding	No action
	No response	Response					
Africa	1	3	23	5	14	3	1
Americas	0	2	23	10	3	1	9
Arab States	1	1	8	4	0	0	4
Europe and Central Asia	0	3	20	9	4	2	5
Asia and the Pacific	1	3	33	12	10	10	1
Subtotal	3	12	107	40	31	16	20
Standards	0	5	66	13	38	11	4
Employment	1	7	57	28	14	6	9
Social protection	0	1	13	9	4	0	0
Social dialogue	0	0	2	14	13	1	0
Other (GENDER)	0	1	8	7	1	0	0
Subtotal	1	16	158	70	58	17	13
Total	4	28	265	110	89	33	33
Percentage	–	–	–	42%	34%	12%	12%

28. The review of the 2011 exercise on follow up shows a slight improvement over the previous year with 76 per cent of recommendations reported as completed or partially completed, compared with 72 per cent in 2010. Non-response fell by 8 per cent. EVAL is eager to work with departments and sectors to make the process easier to understand and more useful. The benefits arising from follow-up to recommendations, especially with regard to the participation of governments and employers' and workers' organizations, are derived from the idea that the knowledge gained from evaluation is absorbed by management in a sustainable manner. In 2013, EVAL will aim for even better compliance by reviewing the quality of recommendations and efforts to increase the participation of management in effective, efficient and relevant follow-up reporting.

> *Biennial milestone 2.1 (2012–13): At least a 25 per cent improvement in reported use of evaluations by constituents over 2011 levels.*

29. An analysis of the involvement of national constituents in the follow-up to evaluations was carried out for the period 2010–11. A total of 629 follow-up actions to recommendations were reviewed, of which 65 required services to be delivered to constituents, their participation in activities, or consultations with them. Of these, roughly 75 per cent were aimed at all constituents, 20 per cent at governments, and the rest at either employers' or workers' organizations only. The follow-up reports also indicated that constituents' involvement in the follow-up was high or moderate for 86 per cent of the 65 recommendations, as depicted in figure 3 below. This data will serve as the baseline for the 2013 milestone on their reported use of evaluations.

Figure 3.   Constituent involvement in response to evaluation
recommendations, 2010–11

<div style="border:1px solid #000; padding:8px">
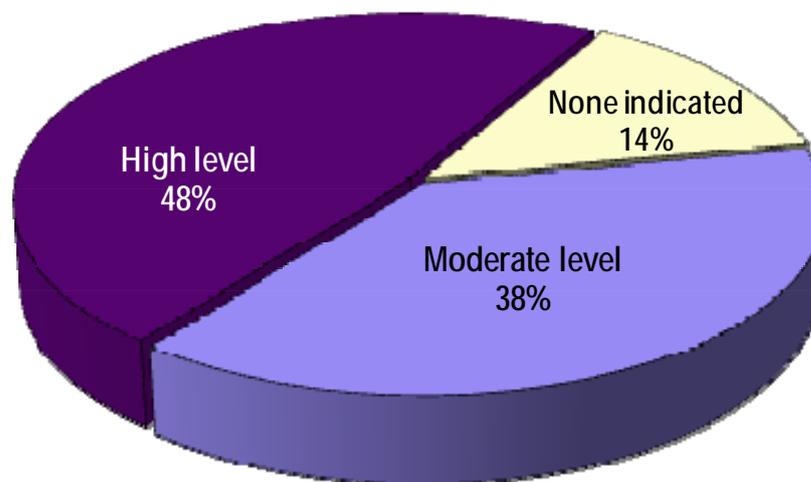
Biennial milestone 2.2 (2012–13): All internal and self-evaluations accessible and searchable in the database.

</div>

30. In 2011, the Office raised the threshold for independent evaluations to $1 million and formalized the policy for internal and self-evaluations. In addition, EVAL, together with the Partnerships and Development Cooperation (PARDEV), made the collection of internal and self-evaluations a priority. Table 4 shows the increase in internal and self-evaluation reports submitted to EVAL since 2007. Additionally, the regions conducted nine Decent Work Country Programme reviews in 2011, a dramatic increase compared to previous years.

Table 4.   Internal and self-evaluations submitted to EVAL, 2007–11

Internal and self-evaluations	Year					Total
	2007	2008	2009	2010	2011	
Decent Work Country Programme reviews	4	2	1	4	9	20
Internal and self-evaluation reports from technical cooperation projects	21	2	4	7	24	58

Outcome 3: Evaluation capability expanded through enhanced knowledge, skills and tools

31. Since 2007, the ILO has supported an evaluation network involving focal persons in regions and technical sectors to ensure the quality and credibility of decentralized evaluations. Each biennium, EVAL convenes the members of this network, together with the Bureau of Programming and Management (PROGRAM) and PARDEV, to take stock of the evaluation efforts undertaken during the previous biennium; to exchange ideas and

brainstorm on developing the workplan; and to identify and discuss bottlenecks and challenges faced by the regions and sectors. In December 2011, the meeting flagged, among other issues, the need to strengthen management's accountability for follow-up to evaluations, including internal ones, within regions and sectors. Related to this is the need to check the quality of recommendations for relevance, specificity and potential for actionable response. Additional priorities included further developing training materials and tools, improving efforts by field staff to involve governments and employers' and workers' organizations in follow-up, and improving the Office's adherence to evaluation standards and good practice.

32. As the result of a joint EVAL–PARDEV initiative, the evaluation network is also benefiting from improved workflow and reporting procedures for technical cooperation projects including: (i) procedures to consistently identify evaluation plans and action at the design and approval stages; (ii) a clear workflow for the identification of evaluation managers, consultant management and approval procedures; (iii) a process to ensure that constituent involvement and management response is routinely considered at all key stages of the evaluation process; and (iv) the announcement of the routine collection of internal and self-evaluations, which is required by the ILO's evaluation strategy 2010–15, as a priority.

> *Biennial milestone 3.1 (2012–13): 75 constituents and 75 ILO staff trained during the biennium.*

33. **National evaluation capacity development**: EVAL commissioned a study to examine the participation of governments and employers' and workers' organizations in national evaluation systems. It covered five countries and placed a special emphasis on the existing priorities of international development organizations and line ministries aimed at promoting the participation of civil society, and workers' and employers' organizations in national evaluation systems. The study found that ILO country offices lagged slightly behind other United Nations agencies in awareness of national evaluation capacities and systems. The majority of country offices felt that national evaluation capacity development should be part of the ILO's country programme activities, while a minority believed it should be included only when requested by the national partners. The priorities of ILO country offices were to measure progress towards achieving Decent Work Country Programme goals and to achieve consensus among the partners for these goals. The offices also organized a range of capacity-building activities for them.

34. While it was co-chair of the Evaluation Capacity Development Task Force of the United Nations Evaluation Group (UNEG) in 2011, the ILO, in collaboration with the United Nations Children's Fund (UNICEF), UN Women and the United Nations Development Programme (UNDP), worked towards developing a UNEG guidance tool on how to strengthen national evaluation capacity systems. The tool is not meant to be prescriptive but is intended to highlight key elements to consider when working on national evaluation capacity development. The developers of the guidance tool benefited from the study on the role of governments and employers' and workers' organizations in national evaluation systems.

35. **Constituent training**: EVAL recently completed a CD-ROM in three languages (English, French and Spanish) of all the training materials needed to increase the evaluation capacity of ILO tripartite constituency. It contains 17 modules, each with a description of the session, slide presentations, and learning activities. EVAL has agreed with the Bureau for Workers' Activities (ACTRAV) and the Bureau for Employers' Activities (ACT/EMP) to work with worker and employer specialists in order to adapt and integrate selected modules from the CD-ROM into the various management and/or governance training

programmes that are being offered through the ILO International Training Centre (ITC) and in the regions.

36. **Evaluation manager certification**: EVAL has agreed with the ITC and the Human Resources Development Department (HRD) to launch a staff development programme intended to certify selected ILO staff to become evaluation managers. The certification programme will consist of a three-day workshop, possibly with a guided practical component. EVAL will partner with the ITC on all three components of the training strategy.

37. **Constituents and ILO officials trained in evaluation in 2011–12**: In addition to formal training, the ILO regions carried out evaluation capacity-building activities targeting both staff and constituents. Table 5 compiles data on the numbers participating in capacity-building activities by region. As for 2012, the Office has already met the biennial milestone within a single year.

Table 5.   ILO evaluation capacity development – Numbers of ILO staff
and constituents benefiting, by location – 2012 *

Persons trained	Africa	Americas	Arab States	Asia and the Pacific	Europe	HQ	Total
ILO staff	9	11	35	15	12	0	82
ILO constituents	32	0	20	17	13	0	82
Total	51	11	55	32	25	0	164

* Any training that is less than one day is counted as sensitization.

> *Biennial milestone 3.2 (2012–13): Internal governance document on evaluation network: Approach, roles and responsibilities adopted and applied.*

38. The new evaluation strategy identified the importance of establishing "evaluation officer" as a generic job description and of improving the guidance on their roles and responsibilities. The draft job descriptions for P2–P5 evaluation officers prepared by EVAL have now been submitted to HRD for further action. Work on an internal governance document for the evaluation network will start in 2013. This work depends on the consolidation of the regional evaluation focal posts and progress as regards the evaluation manager certification training aimed at the informal network of more than 90 ILO officials managing evaluations each year.

H.   Improving the use of evaluation knowledge systems

39. The key component of the ILO evaluation knowledge system is the i-track database used to store information on evaluations planned, in process or completed. The knowledge base supported through this system has continued to expand and feed into other ILO knowledge systems, including the donor dashboard. Table 6 maps the steady expansion of products stored and accessible to all ILO staff since 2005.

Table 6.   Overview of the volume of evaluation knowledge reports
             captured and accessible, 2005–11

Type of knowledge/report	Year							System total
	2005	2006	2007	2008	2009	2010	2011	
Independent evaluations – ILO managed – annual reporting	65	55	42	66	63	71	95	457
Independent evaluations – externally managed – joint evaluations	–	–	–	2	5	12	54	73
Independent evaluations – EVAL managed – country programme, thematic and strategy	1	3	3	4	3	6	4	24
Meta-evaluations/internal reviews – EVAL managed	–	–	–	–	1	2	2	5
Independent evaluation summaries – ILO public website	–	25	28	37	177	75	95	437
Information CDs / training CDs	–	–	–	–	1	1	2	4
Independent evaluation reports on the donor dashboard	–	–	–	–	–	391	84	475
Internal evaluations	–	8	9	6	8	12	26	69
Decent Work Country Programme internal reviews	–	1	3	2	1	4	9	20
Research think pieces	–	–	–	–	–	–	4	4
New guidelines and guidance notes	–	5	8	0	1	3	35	52
Newsletters	–	–	–	–	–	3	3	6
Lessons learned	–	–	–	115	340	121	116	692
Recommendations	–	–	–	630	717	642	325	2 314

40. **Organizational learning and policy-focused evaluation**: In addition to systems for storing and making accessible evaluation-generated knowledge, the strategy calls on EVAL to contribute to organizational learning through the effective generation of lessons learned. Over the past few years, the ILO has launched a number of initiatives aimed at reinforcing its evaluation function with a view to supporting more systematic institutional learning. However, the general perception is that lessons generated by evaluations are not yet adequately incorporated into the Office's programme and policy decisions. Therefore, EVAL undertook a study to explore both how and to what extent greater policy-based learning can be achieved from project and country programme evaluations. The study found that project and country programme evaluations are not yet geared towards identifying lessons on national decent work policies. This may, in part, reflect the treatment of policy in the results frameworks of many projects, which tend to be more operational and to focus on action on the ground. Suggestions in the study ranged from simply encouraging evaluators to place more emphasis on the policy rationale of the projects when they evaluate them, to the more radical idea of ensuring that all projects/programmes are designed with reference to their contribution to either building capacity for policy implementation or, alternatively, to promoting policy change.

41. Since 2009, EVAL has been extracting lessons learned and coding them in the i-track database. A review of lessons learned reveals that evaluators tend to class almost 50 per cent of all lessons learned as process or administrative lessons. To correct this in the coming year, EVAL will alter its guidance with a view to improving the usefulness of lessons learned by requiring evaluators to concentrate more on policy relevance and technical know-how aspects. EVAL intends to increase the quality of these lessons learned

within thematic contexts by establishing validation exercises and discussions on specific technical areas of the ILO's work.

Part II. Assessing the ILO's effectiveness and results

A. Introduction

42. The revised 2011 ILO strategy for evaluation requires evaluative efforts to synthesize information on the Office's overall effectiveness and achievement of results. Last year's report presented the findings of three major studies. The discussion of these findings during the November 2011 session of the Governing Body resulted in the formulation of a plan of action for the implementation of the recommendations and suggestions endorsed during the March 2012 session of the Governing Body. Appendix I provides an update on the status of implementation of this action plan.

43. This year, the performance section looks at results-based management practices by distilling some of the generic lessons drawn from the high-level evaluations conducted during 2012. This section also reports on the results from a review of the effectiveness of quality appraisal systems introduced by the Office for new Decent Work Country Programmes and technical cooperation projects, and of compliance with new performance reporting systems. Finally, the section outlines a review of performance rating practices in the ILO and how they can be improved.

B. Results-based management and ILO effectiveness: Insights from high-level evaluations

44. In preparation for the next Strategic Policy Framework 2016–21, last year EVAL undertook an extensive review of the Strategic Policy Framework 2010. The review showed substantial progress towards implementation and compliance with the results-based management policy in the ILO but also identified areas for improvement, including the need to better align resource allocations with results; the importance of greater coherence between results and the accountability framework; and the need to improve the quality of indicators, assumptions and risk assessments in the Strategic Policy Framework. This year, EVAL reviewed the findings of two strategic and one Decent Work Country Programme evaluation undertaken in 2012 in order to gain an insight into the three aspects outlined in the annual evaluation report for 2010–11.

45. **Accountability for timely delivery of outputs, good quality**: The evaluations found that the quality of work and delivery of outputs were on track and that managers were motivated to achieve performance targets. However, primarily at the planning stages, the system itself generates competition among managers when designating targets and securing resources to deliver the outcome-based workplans and underlying country programme outcome initiatives. The lesson learned is that the current programming process for outcome-based workplans and country programme outcomes functions reasonably well in terms of accountability for output delivery but that competition may compromise effective prioritization.

46. The Office has made significant advances in aligning the Decent Work Country Programme with results-based management principles but the scope for evaluating the results frameworks and monitoring and evaluation plans remains problematic. The most

prevalent issues are the uneven use of baselines, measurable indicators, milestones and risk assumptions at the country programme outcome stage.

47. **Accountability for relevance, validity, sustainability and attainability of the assumptions establishing the link between outputs and expected outcomes**: In the various strategies, assumptions are most critical at the country programme outcome stage, since risks are highest in countries where the ILO's influence is subject to national, political and economic factors. In the strategy evaluations, overall, adequate management of risks associated with assumptions was found in all areas except for sustainability, where the ILO's assumption that national implementation would follow as a result of ILO support was not always well founded.

48. Although the preparation of outcome-based workplans was intended to promote better collaboration, coordination and coherence, the one-to-one linkage of country programme outcomes and programme and budget outcomes has posed some difficulties for programmes that are more cross-cutting in nature in terms of devising coherent and synergetic workplans embedded in the country programme outcome and outcome-based workplan system.

49. **Accountability for the quality of the operational risk management**: Operational performance was found to be adequate due to good coordination and communication between field and headquarters. However, risk management challenges were found to be greatest for resource risks linked to longer term funding, and for operational risks related to the ambiguity of roles and responsibilities, such as that between the field and headquarters or within line management structures.

As the Office is currently in the process of reviewing the structure of the programme and budget as part of the 2014–15 proposals, in response to last year's recommendations, this year, no additional ones under this heading are included.

C. The ILO's evolving quality assurance

50. The annual evaluation report 2010–11 outlined results from several studies highlighting weaknesses in the ILO's performance in terms of designing, monitoring and reporting on projects and Decent Work Country Programmes. As a follow-up to this exercise, EVAL reviewed the use and usefulness of several mechanisms introduced by the Office over the past few years in order to improve quality in these areas. The three tools reviewed are the project appraisal system, the Decent Work Country Programme quality assurance mechanism and a standardized tool and template for reporting project implementation performance.

51. Given that the appraisal tools themselves have undergone quality checks, EVAL focuses its analysis firstly on the actual use of the two appraisal types and on the performance reporting template, and secondly, when these are used, on the evidence available to indicate their effectiveness. The preliminary results of these reviews will contribute to the planned high-level evaluation of the ILO technical cooperation strategy currently proposed for 2013.

D. Quality appraisals of technical cooperation projects

52. The ILO appraisal mechanism consists of an analytical assessment of the design of technical cooperation programmes and project proposals. It was established in June 2010.

Its purpose is to ensure that technical design standards have been met, results-based management has been applied, and that the proposals contribute to achieving the ILO's priorities and Decent Work Country Programmes.

53. PARDEV records indicate that 210 technical cooperation projects were submitted for appraisal from January 2011 to June 2012. Of those with budgets exceeding US$1 million, a random sample of 20 per cent was taken, after which a final set of 15 were selected on the basis of the full documentation available to complete this analysis.

54. PARDEV has developed three reporting templates: concept note appraisal, initial appraisal and final appraisal report. The final appraisal report authorizes its submission to the donor. As required, appraisals of project proposals are completed by different parts of the Office but the procedures followed differ from region to region. In the case of Asia, the Regional Office has provided initial comments aimed at improving quality and forwarded the proposal only when quality standards were met. In other cases, headquarters and the regional offices have provided feedback simultaneously. In addition, various parts of the Office are asked to appraise mainly those sections of the proposal that fall within their area of expertise.

55. Compliance was measured according to the quality and timeliness of the feedback provided by the various units that reviewed proposals on the basis of their adherence to PARDEV's prescribed template (appraisal checklist for technical cooperation proposals) and response time. The effectiveness of the appraisal system was analysed on the basis of notable improvements in the quality of the proposals.

Compliance

56. Overall, PARDEV and other contributors complied with the appraisal process. The draft proposals were found to have been widely circulated among the relevant ILO staff, who could then communicate and coordinate with those who had made the proposal both within and outside the formal appraisal process. This is considered to be a major improvement over the previous closed process for developing proposals.

57. As regards the feedback provided, PARDEV comments were available for 93 per cent of the sampled technical cooperation projects. Not all of this feedback adhered to the standard appraisal checklist and some comments may have been sent in text form. The comments made by PARDEV were found to be very comprehensive although they only concentrated on a few sections of the draft proposals. Useful comments were provided to the units that had made the proposals with a view to aligning the project objectives and strategies with donor priorities and the ILO's ongoing efforts. In terms of appraisal criteria, relevance and strategy fit, programme logic, knowledge management and sharing, and presentation were found to have received maximum attention. These criteria are consistent with appraisal priorities, given that one of the main objectives of the appraisal is to ensure that technical cooperation projects provide measurable contributions to ILO outcome-based work planning.

58. As has been mentioned, many of the technical cooperation projects were appraised at earlier stages, (e.g. four of the 15 technical cooperation proposals were fully appraised by the Asia and the Pacific region and revised prior to their review by PARDEV). The response time for appraisal feedback was found to be within the prescribed time frame (five days) for the sample cases. In fact, nearly 40 per cent of the sampled technical cooperation projects were appraised by PARDEV within two days of them receiving the drafts.

Effectiveness

59. The assessment of effectiveness was carried out based on a four-point scale where 1 is highly unsatisfactory, 2 is unsatisfactory, 3 is satisfactory, and 4 is highly satisfactory. The appraisal criteria described in the technical cooperation appraisal checklist were used for scoring the proposals before and after appraisal. This was done in order to assess the degree to which the comments provided are used to improve the quality of proposals. Since appraisal templates were not always followed, the comments made by PARDEV were used as the basis for scoring the pre-appraisal quality of the sample. Criteria not mentioned by PARDEV in its appraisal were assumed to be satisfactory or not requiring significant change. (This methodology did not measure improvements made prior to PARDEV's appraisal.)

60. Figure 4 below shows the results of EVAL's assessment of changes in the quality of proposals before and after feedback from PARDEV following an initial round of feedback and revisions in at least some of the regions. Notable improvements were observed in the area of the results framework; followed by that of relevance and strategy fit; and in that of presentation, which PARDEV had identified as weaknesses in many proposals. While this was not captured in the matrix, feedback was also regularly provided on and applied to the areas of budgeting compliance, handling of procurement, and evaluation requirements.

Figure 4.   Average quality ratings of technical cooperation proposals, pre- and post-appraisal
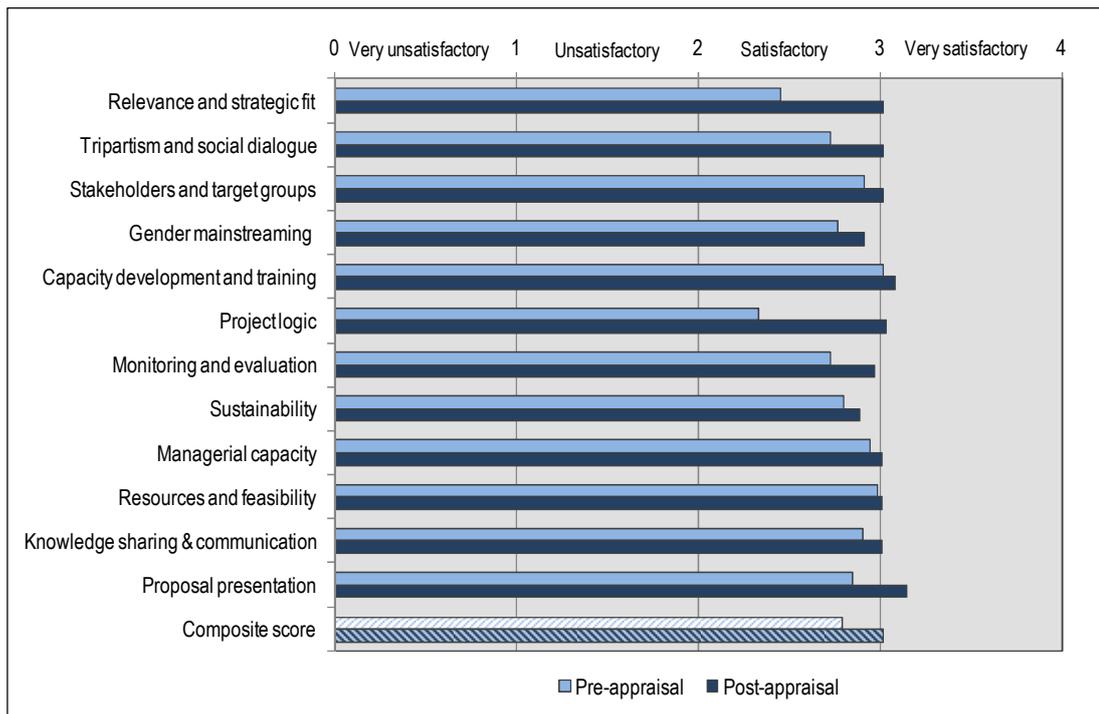
61. As figure 4 suggests, the results show that the appraisal process is contributing to and improving project design through targeted and systematic feedback from the relevant ILO departments and units on critical proposal areas. The results suggest that the quality of the proposals reaching PARDEV has already improved as a result of the review process. The magnitude of improvement shown in the graph suggests that the PARDEV appraisal feedback has further improved quality in the particularly weak areas of project logic and strategic fit, thereby ensuring that proposals meet the required standard of satisfactory on all criteria.

62. Another interpretation of the results may suggest that those who made the proposals are not making effective use of the feedback they receive to improve proposal quality. Even post-appraisal project proposals are still clustered around the satisfactory quality level with none achieving a composite post-appraisal quality rating above 3.5. The results suggest that appraisal processes alone cannot improve proposal quality beyond a certain level, given the limited time frame and the advanced stage of proposal development. Instead, the combination of guidance, training and technical support currently provided may need to be enhanced and introduced as early as possible to complement the appraisal tool.

Recommendations

63. **The appraisal function has been well established at headquarters. However, regional capacities need strengthening in order to fully roll out this function.**

64. **The Development Cooperation Branch (CODEV) should continue to strengthen the linkages between its supervisory and oversight role and its guidance and capacity-building work aimed at improving the quality of project design during the proposal stage. This may involve targeted support earlier in the proposal development stages.**

65. **The Office should consider stronger mechanisms for linking final proposal quality to the accountability of the unit from which the proposal originated. Where quality is found to be low, plans for follow-up at the post-approval stage should become more systematic.**

E. Decent Work Country Programme quality assurance mechanism

66. The assessment of the Decent Work Country Programme quality assurance mechanism considered the level of compliance with the ILO Decent Work Country Programme guidelines and the effectiveness of the process on the quality of the approved Decent Work Country Programmes. It focused on the efficient implementation of the quality assurance mechanism in order to significantly reduce the Decent Work Country Programme preparation cycle while maintaining the strategic relevance of the exercise. According to ILO records, 34 Decent Work Country Programmes were appraised over the past two years. Of the nine (selection was also based on the availability of the necessary documentation) analysed for this study, only one made use of the revised quality assurance mechanism guidelines issued in late 2011.

67. The assessment focused on: (i) compliance with the quality assurance mechanism templates; and (ii) the effectiveness of the exercise on the quality of the approved Decent Work Country Programme document.

Compliance

68. Compliance was measured in terms of the use of the ILO quality assurance mechanism appraisal checklist, and the extent to which the appraisers addressed the established criteria and questions. The analysis of compliance by region shows that, in the past two years, the Americas has not initiated the quality assurance mechanism process to improve their draft Decent Work Country Programmes. The timeliness and efficiency of responses could not be fully assessed due to too little information being available on the dates when feedback was sent.

69. In the nine Decent Work Country Programmes reviewed, regional offices, PROGRAM, PARDEV and EVAL completed just under half of the requested appraisals, and the Bureau for Gender Equality (GENDER), just over 50 per cent (see figure 5 below). As with project appraisals, many of those completing the appraisal respond to those questions related to their technical area and leave blank the sections with which they are less familiar or in which they have less expertise. Under the revised system, the quality assurance mechanism continues to require inputs from regional offices, PROGRAM, PARDEV, EVAL, ACT/EMP, ACTRAV and GENDER.

Figure 5. Availability of quality assurance mechanism appraisal records by appraiser group
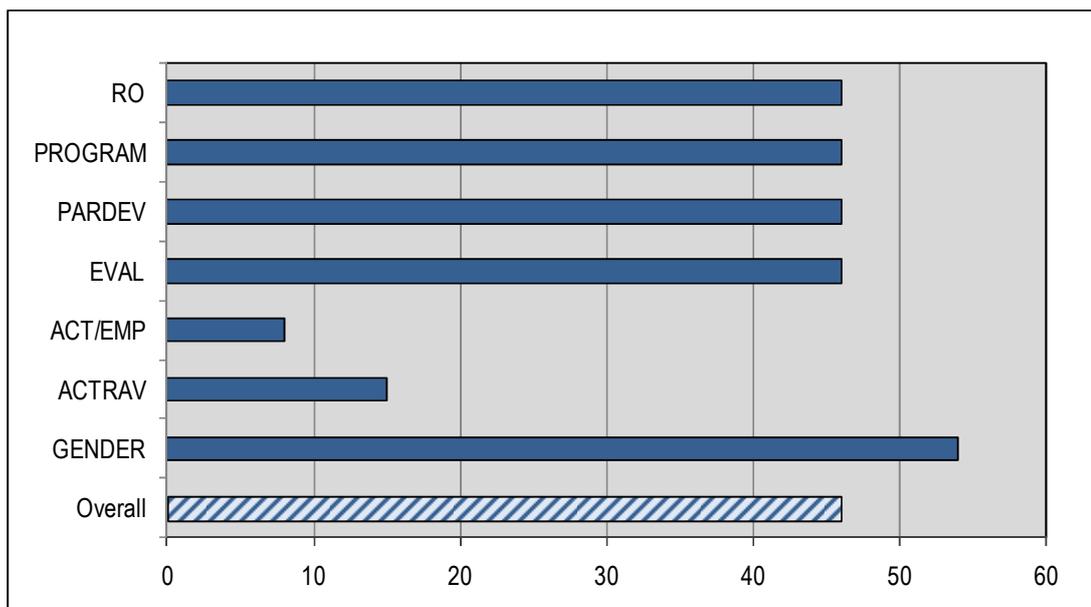
Effectiveness

70. The effectiveness of the quality assurance mechanism was assessed on the basis of the comprehensiveness and quality of the feedback provided and the degree to which this resulted in improvements to the Decent Work Country Programme documents. This involved mapping the key observations and analysing the precision and clarity of feedback, whether appraisers made a specific reference to sections that needed improvement, and whether alternatives were provided to fill the gaps in those areas.

71. The assessment indicates mixed results in terms of the effectiveness of the process on improving the quality and enhancing the results orientation of Decent Work Country Programmes. A key finding is that closer involvement of regional offices in Decent Work Country Programme development and appraisal helps to improve the document. This is apparent from the scoring results for Africa and the Asia and the Pacific regions. Feedback from these regional offices was found to be very comprehensive, constructive and more in step with country contexts. This finding strongly supports the revised quality assurance mechanism procedure requiring regional offices to take the lead in the process.

72. Overall, post-quality assurance mechanism Decent Work Country Programmes showed improvements in the areas of strategic fit, programme logic, presentation, knowledge management and resource mobilization, as depicted in figure 6 below.

Figure 6.   Average quality scores of Decent Work Country Programmes
quality assurance mechanism criteria, pre- and post-appraisal
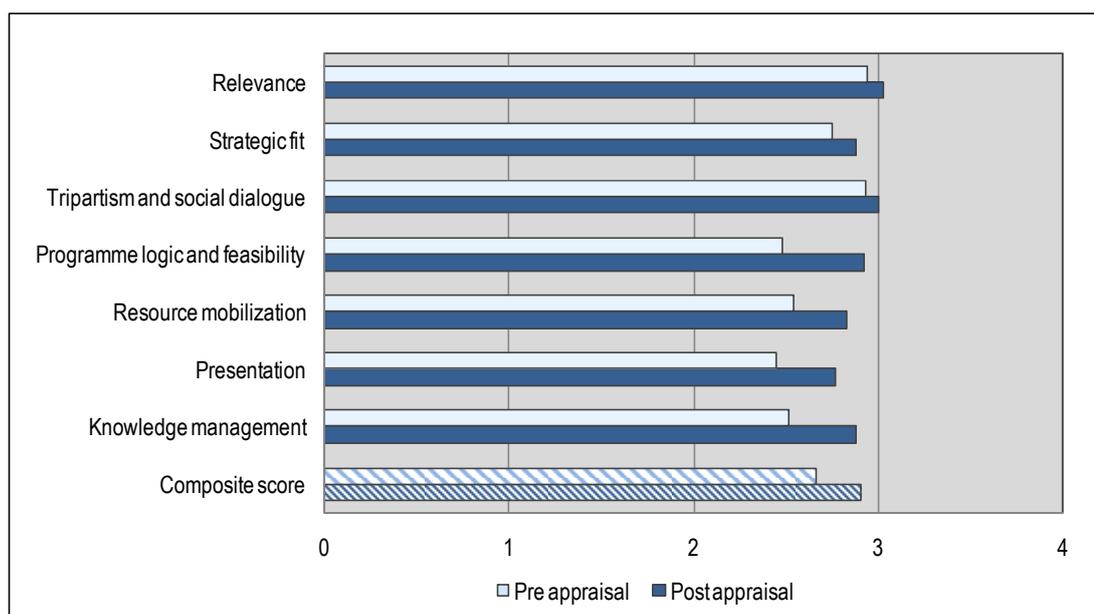
73. Most appraisers provided feedback that was clear and concise. In most cases, comments made specific references to the areas requiring improvement and also provided useful alternatives.

74. The results frameworks received the most frequent and elaborate feedback from appraisers, which resulted in greater improvements. As shown in figure 7 below, the quality of final Decent Work Country Programme documents was found to be generally satisfactory, with one being assessed as highly satisfactory.

Figure 7.   Composite scores of Decent Work Country Programme quality
assurance mechanisms by country, pre- and post-appraisal
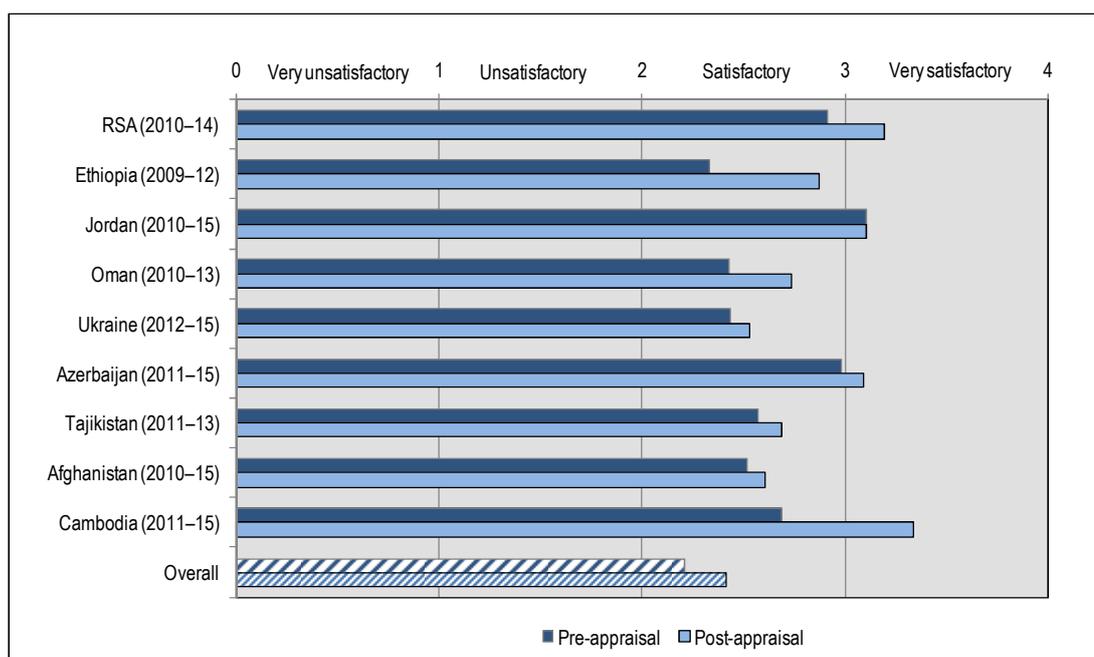
Key lessons

75. The availability of quality assurance mechanism records poses a major challenge in assessing the degree of compliance. While records indicate that substantial efforts are made by the country offices and appraisers to enhance the quality of Decent Work Country Programmes, there is not enough evidence to ascertain the degree to which the appraisal comments are taken into account when finalizing the Decent Work Country Programmes.

76. There is no single repository of quality assurance mechanism records for Decent Work Country Programmes. Moreover, the time taken at each stage of the appraisal is not recorded, which makes time-efficiency impossible to establish. At times, appraisers do not record their designation or location, making it difficult to identify the appraising unit.

77. The new quality assurance mechanism checklist limits feedback to appraisal criteria that are relevant to specific units. This makes the entire appraisal cycle more efficient, the process easier for the country offices and less time-consuming for appraisers. This is an improvement on the pre-2011 template.

Recommendations

■ **Standard guidelines for appraisers that explain the factors to be kept in mind while appraising should be developed. On the whole, apart from highlighting weak areas, providing specific alternatives for improvement should be encouraged. The appraisal could also be used to identify the areas for capacity building in country offices.**

■ **Country offices should be encouraged to file their response to appraisal comments, which should be part of the quality assurance mechanism repository for each Decent Work Country Programme. The Office should maintain repositories for documents relating to quality assurance mechanism appraisals, drafts and final Decent Work Country Programmes at headquarters or in regional offices.**

F.   Progress reporting of project performance

78. In 2010, the ILO revamped its project implementation monitoring tools and introduced a systematic approach to tracking the achievement of project outputs and outcomes relative to their estimated time and cost parameters through the technical cooperation progress reports template. These are critical tools for forward-looking strategic positioning, organizational learning and sound management. They incorporate a greater scope for evaluation, and capture lessons learned and emerging good practices.

79. In view of differences in the reporting requirements, not all progress reports use the new template. However, PARDEV actively encourages donors to accept the new system as the revised reporting format for interim and final progress reporting. The first self-evaluations using these new formats were available in 2012.

80. The purpose of the assessment is twofold: to determine the level of compliance with the new standard progress reporting system, and the effectiveness of these reports as project management and organizational learning tools. A sample of 16 technical cooperation progress reports was used (one completed in December 2010, three completed in 2011, and 12 covering the period between 1 January and 30 June 2012). The sample selection was based on the following criteria: (i) regional representation; (ii) diversity in terms of

reporting criteria (i.e. ILO mandated and/or donor required); and (iii) reports for operations exceeding US$1 million submitted to PARDEV.

Compliance

81. The availability of technical cooperation progress reports was a major constraint to conducting this analysis, in part due to there being no single repository for progress reports. Instead, each region and sector is responsible for managing and storing technical cooperation progress reports that fall under their area of responsibility.

82. The Asia and the Pacific region was found to have the most systematic process for storing technical cooperation progress reports due to its knowledge sharing platform. A similar system is under construction by PARDEV for use at the ILO global level.

Effectiveness

83. The results of the quality assessment of sample progress reports show that on average the quality of most reporting sections was found to be unsatisfactory. No regional office or sector was found to systematically check progress reports for quality and it is unclear how the reports had been fed into the broader programme and budget programming processes.

Figure 8. Average scores of technical cooperation progress report quality, by reporting category

Recommendations

■ **Systematic quality assurance of technical cooperation progress reports should be conducted by the responsible administrative units in the regions and headquarters, with oversight exercised by PARDEV.**

■ **Technical cooperation progress reports should inform decision-making and provide input for PARDEV's annual reports on the overall implementation of the ILO's technical cooperation portfolio.**

■ **In the absence of an all-encompassing monitoring and evaluation system, the Office should establish a centrally managed knowledge exchange system where**

technical cooperation progress reports can be stored and accessed by all internal stakeholders.

■ **ILO donors should support, to the extent feasible, the use of the technical cooperation progress report approach to progress reporting.**

G. Evaluation of Regular Budget Supplementary Account activities

84. The ILO established the RBSA to support the Decent Work Agenda through flexible non-earmarked voluntary contributions to the ILO work programme. RBSA is a key resource in delivering decent work results. The intent of this section of the report is to continue with the practice established in last year's annual evaluation report to report on evaluations of RBSA-funded initiatives and on the effective use of RBSA funds set aside for monitoring and evaluation.

85. The Office completed six evaluations of programmatic work with a sizeable RBSA-funded component. Of these, four came from Africa and two from Latin America. Five were independent evaluations and one was internal. Most covered multi-year work programmes for the period from 2009 to 2011. Most RBSA allocations for this period fell under the results framework of multiple Decent Work Country Programmes so that the scope of the evaluations was mainly thematic.

86. EVAL conducted a meta-analysis of the ILO's performance based on the findings of six evaluation reports. The analysis criteria included: (i) relevance; (ii) coherence; (iii) effectiveness; (iv) efficiency; (v) sustainability and impact; and (vi) monitoring and evaluation framework and evidence of results.

Key findings

87. All sample evaluations followed the framework for evaluation recommended by UNEG and also provided useful recommendations and lessons learned. Findings relate to the six reviewed evaluation reports and have to be considered in conjunction with other reports on the overall effectiveness of RBSA-funded initiatives. Most RBSA interventions (five of the six) were found to have added value in achieving the regional strategy and Decent Work Country Programme outcomes they supported. Table 7 below lists the evaluations reviewed for this analysis. The assessment results are shown in figure 9.

Table 7. Evaluations of RBSA-funded initiatives reviewed for the assessment (US$)

Allocation title	Allocation amount
Support for the development of public policies for the promotion of cooperatives in the Plurinational State of Bolivia	200 000
Social security quantitative training for Africa (QUATRAIN AFRICA)	3 948 279
Extending social security to African migrant workers and their families	2 866 710
Youth Employment Support – Jobs for the Unemployed and Marginalized young People (YES–JUMP)	1 299 274
Improving institutional capacity to govern labour migration in North and West Africa	3 947 368
Allocations of RBSA 2010–11 funds in support of workers' and employers' organizations	800 000
Grand total	13 061 631

Figure 9.   Average performance ratings of RBSA-funded initiatives, by evaluation criteria
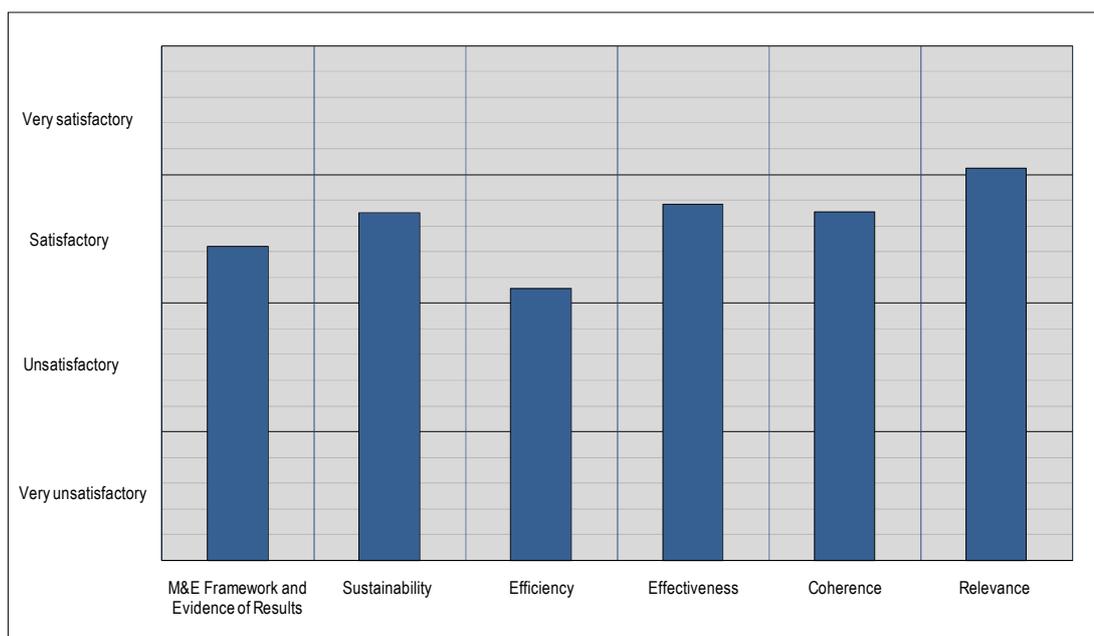
88. The evaluations pointed out that the RBSA-funded interventions supported country programme outcomes but often were not linked to a clear logical framework and monitoring system, largely due to the uneven use of a monitoring system for the country programme outcome to which the RBSA was linked. Most evaluations found that while the purpose, outcomes and outputs of country programme outcomes were clearly defined, measurable indicators and milestones for tracking progress were far less evident. Exceptions were RBSA allocations that supplemented initiatives funded through traditional technical cooperation projects, for which logical frameworks and monitoring systems were available and aligned with the relevant country programme outcome.

89. Several evaluations found that more frequent performance reporting covering a wider range of criteria was needed to ensure an effective oversight of the country programme outcome and to identify lessons learned. This is strongly encouraged in the revised Decent Work Country Programme guidelines but is not compulsory.

90. Roughly half the evaluations noted that in cases involving the collaboration of different ILO offices (country offices, regional offices and headquarters) to deliver the RBSA-supported country programme outcomes, there were challenges related to ambiguity over roles and responsibilities. As the ILO makes advances towards allocating RBSA for country programme outcomes, it will need to provide staff with guidance on how to merge practices previously established for either regular budget programming or technical cooperation programming.

Recommendations

■ **Considering that country programme outcomes are the ILO's main results framework for activities at the country level, linking RBSA allocations to country programme outcomes through a results-focused monitoring system needs to become more consistent.**

■ **EVAL should continue to monitor and assess country programme outcome monitoring and evaluation practices and report to the Governing Body on the value added to country programme outcomes through RBSA investments.**

H.   Performance ratings in ILO evaluations

91. Since 2006, the ILO evaluation function has used ratings in evaluations as an added dimension to communicating evaluation findings. However, rating systems can be misleading if the quantitative values applied to ratings are not used consistently (reliability) or accurately (validity). For this reason EVAL contracted Professor William Trochim of Cornell University to review its rating methodologies to find ways to strengthen and improve them.

92. Being a proponent of ratings as a means of reinforcing results-based management systems, the reviewer endorsed the ILO's move to apply ratings as part of evaluation methodologies. If used appropriately, ratings can be relatively easy to apply and can summarize a broad range of key features in just a few numbers.

93. Overall, the review offered several constructive suggestions to improve the various tools, such as more systematically verifying consistency between the descriptive findings and the quantitative scores given. "Wherever possible and appropriate, evaluators should be encouraged to provide multiple forms of evidence (triangulation) to justify the award of a particular score." [2]

94. A second area for improvement would be the calibration of scores to convey similar performance criteria and levels, which could involve independent checks of score choices. The review advised the ILO to check for overlap between indicators. Finally, the review cautioned against aggregating ratings into a composite performance score for a project or programme or subsuming multiple projects under the programme level. Table 8 summarizes some of the feedback contained in the report on each tool.

Table 8.   Summary of findings for ILO rating methodologies

	Weighted scoring to indicate whether topic can support an independent evaluation
Evaluability assessment	Complements descriptive analysis to signal through scores the preparedness of the topic for evaluation
	Need to better distinguish between ratings that reflect poor performance versus prematurity of timing
	Independent evaluator rates performance aspects of DWCP and individual CPO
DWCP review	Ratings validated through constituent and staff feedback
	Link between findings and ratings not always well supported in reports

[2] A summary of the study is available on request from eval@ilo.org.

Strategy evaluation	*Qualitative scoring to indicate strategy performance within evaluation criteria*
	Alignment with SPF and P&B outcomes and indicators
	Need to explain purpose of scores within evaluation methodology
Project performance rating	*Ex-post rating of project performance using standard criteria and indicators based on final independent evaluations*
	Matrix includes qualitative description of each performance level or score on a four-point scale, offering a degree of objectivity to the rating process
	Indicators should be reviewed to eliminate overlap and tighten calibration

95. In the follow-up to the review, EVAL will revise its rating practices based on **recommendations** by taking the following actions:

■ Being consistent in terms of the choice of scales and criteria, ideally moving to a six-point scale and applying OECD/DAC evaluation criteria as the default performance matrix for evaluations.

■ Avoiding the use of aggregation and weighting of data within an individual assessment.

■ Integrating the management information system for compiling and storing evaluation-based performance data across all tools and time periods.

■ Carrying out a peer review of all ratings used in high-level evaluations prior to their finalization.

Draft decision

96. *The Governing Body takes note of the present report and endorses the recommendations to be included in the attached rolling plan of action for the implementation of recommendations and suggestions to be reported on in the annual evaluation report of 2012–13, and confirms the priorities identified in the report on the work programme for 2012–13.*

Appendix I

Plan of action for the implementation of recommendations and suggestions contained in the annual evaluation report

Recommendations AER 2010–11

Suggestions and next steps	Long-term improvements	Short-term actions 2012–13	Who/additional cost 2012–13	Status
1. Assessing the effectiveness of the ILO's strategic framework				
1.1. Operational alignment and resource allocation should be based on the assessment of the four strategic goals' results rather than the reverse. Stocktaking and, potentially in the longer run, impact assessment should be part of the recurrent discussion reports. Their discussion by the ILC required under the ILO Declaration on Social Justice for a Fair Globalization could provide a good basis for this process.	The evaluability of the new SPF results frameworks (2016–20) improved to allow for a sound assessment of the four strategic goals' results.	■ Conduct each year two high-level in-depth strategy evaluations of SPF outcomes or ILC-related topics, as mandated by the Governing Body.	EVAL/none	On track.
		■ Perform an evaluability assessment of the SPF results framework (by November 2012). EVAL produces a comparative analysis of the evaluability of 2010–11 and 2012–13 P&B results frameworks and highlights achievements and areas for improvement in the 2011–12 AER.	EVAL/none	Postponed. The evaluability assessment would be more relevant to the 2014–15 P&B currently under preparation. This would capture adjustments and improvements made to the SPF indicators.
		■ Produce by March 2012 a guidance note on targeted impact evaluation methodologies for technical departments, including guidelines on selection of consultants and quality review of the eventual reports.	EVAL/none	Completed.
		■ To enhance organizational learning, a working paper systematically synthesizing results and lessons learned from completed independent evaluations, impact studies and other research related to the topic under review in the recurrent report will be produced prior to the ILC. EVAL will collaborate with relevant departments to prepare such systematic review as of 2013. Note: The last work item can only be conducted provided the topic of recurrent report is confirmed one year ahead.	EVAL and other departments/ $50,000 per topic	Preparations for a synthesize report based on completed studies and evaluations related to social dialogue have started. The working paper will feed into the recurrent discussions scheduled for 2013.

Recommendations AER 2010–11				
Suggestions and next steps	Long-term improvements	Short-term actions 2012–13	Who/additional cost 2012–13	Status
1.2. Articulate explicitly and codify (systematize) a results-oriented accountability framework for the ILO, differentiated for managers and staff and linked to (weighted) outcome indicators. In the logic of the results framework, managerial accountability for results could be cast in the form of a following triangulation:	The system for supporting implementation of the results framework for the SPF 2016–20 improved to better monitor outcomes including risk management strategies, thus ensuring managerial accountability for results.	■ A critical assessment of the results framework with respect to indicators, measurement criteria and the validity of assumptions will be carried out to inform the development of the next SPF by the end of 2013.	PROGRAM/none	On track. As indicated during the discussion of the 2010–11 implementation report, the Office has ongoing work to improve the qualitative dimension of indicators and measurement criteria. A sample will be submitted as part of the P&B proposals for 2014–15.
(i) accountability for the timely delivery of outputs to the required quality standards, as a necessary condition for achieving the expected results; (ii) accountability for the relevance, validity, sustainability and attainability of the assumptions establishing the link between outputs and expected outcomes; and (iii) accountability for the quality of the "operational" risk management. On the basis of a properly managed triangulation, managers can safely expect the outcomes to materialize.		■ In the context of unit workplans and performance management, establish criteria for the timely delivery of outputs to the required quality standards, as a necessary condition for achieving the expected results.	HRD/PROGRAM/ managers	In May 2012, guidelines on unit workplans were issued. Unit workplans provide a framework for the development of individual performance management plans and provide management with a tool to track progress towards the delivery of outputs. In February 2012, to facilitate the preparation of the beginning of the cycle performance management form for the 2012–13 biennium, a *Guide to Writing SMART Outputs* was issued by HRD and distributed to staff in hard copies. The *Guide* emphasizes the alignment of individual outputs with the unit workplan and provides examples of SMART outputs in support, technical and management areas.
		■ Complete biennial independent validation of the quality of operations and risk management strategies.	PROGRAM with EVAL support/none	Office-wide risk management strategy to be reviewed and redefined in 2013.

Recommendations AER 2010–11				
Suggestions and next steps	Long-term improvements	Short-term actions 2012–13	Who/additional cost 2012–13	Status
1.3. Include the list of critical assumptions in the outcome-based workplan (OBW) template and identify suitable methodological approaches for monitoring the implementation of the P&B and the SPF.	idem.	■ In 2012–13, review assumptions and risks in relation to outcomes in the results framework.	Outcome managers and PROGRAM/none	Information on critical assumptions under each outcome will be included in the OBW template in IRIS and reviewed in 2013.
2. Methodologies for synthesizing evaluation-generated performance information				
2.1. Carry out a meta-analysis of operational performance on a biennial basis, drawing from findings, conclusions, recommendations and lessons learned from independent evaluations completed during the previous biennium.	A sound methodology in place for linking TC performance and results with SPF outcomes.	■ Second biennial outcome-level meta-analysis report completed for November 2013. ■ To better align operational performance data more directly with SPF outcomes, EVAL will fine tune its methodology to better link TC performance with each SPF outcome. Much of the data will focus on country-level indicators, pulling primarily from the list of CPOs associated with each SPF outcome. Note: Action step can only be conducted provided CPO performance data are collected and verified by the Office.	EVAL/$30,000	Planned for 2013. In progress
2.2. Revise the methodology and integrate into evaluation procedures for independent evaluations; collect performance data directly from the independent evaluator(s).	Management information system (MIS) maintained for evaluation data storage and analysis.	■ Beginning in 2012, EVAL will regularly collect data on standardized performance indicators as part of the independent evaluations of TC projects.	EVAL/none	After pilot testing a questionnaire, EVAL considers for validity and reliability reasons that evaluator scoring should be combined with ex-post independent scoring. This approach will be applied for the 2013 report.
2.3. Incorporate in the scope of upcoming evaluations of ILO P&B outcomes consideration of project evaluation findings including performance scoring.		■ Relevant project performance based on the above assessment methodology will be incorporated in the analysis of all 2012–13 high-level strategy evaluations.	EVAL/none	The performance criteria, indicators and ratings were used when assessing the TC portfolios within strategy evaluations conducted by EVAL.

Recommendations AER 2010–11				
Suggestions and next steps	Long-term improvements	Short-term actions 2012–13	Who/additional cost 2012–13	Status

3. Decent Work Country Programmes (DWCPs): Lessons learned

Strengthen the design and implementation of DWCPs based on recommendations contained in meta-study of nine DWCPs in 2010 by ensuring that: (i) the design reflects sound understanding of each country's unique character and flexibility to adapt to external shocks; (ii) constituents are fully engaged in developing DWCPs; (iii) DWCP M&E frameworks are an essential part of the design process and include the necessary resources to implement these; (iv) progress is made towards the institutionalization of the Decent Work Agenda, including its M&E framework, into the country's institutions, budgets and procedures; (v) funding support for DWCP is programme-based rather than project-based, and country office structures are aligned to support the DWCP approach; and (vi) an appropriate communication strategy is developed and funded to share experiences on DWCPs results.	DWCPs are the main vehicle for ILO support and tackle within national development frameworks major decent work deficiencies through efficient programmes that embrace each of the strategic objectives.	■ The DWCP *Guidebook* version 3 (released in December 2011) has addressed all the recommendations contained in the DWCP meta-study reported on in the AER 2010–11. The *Guidebook* reinforces the requirement of direct participation of ILO constituents; contains language suggesting that monitoring plans and implementation plans are mandatory and the quality assurance checklist contains a question on constituents' ownership. During 2012–13, the DWCP *Guidebook* will be actively promoted, used in training events and in other learning and exchange opportunities.	PROGRAM/none	Completed.
		■ Currently the ILO is moving towards less earmarked extra-budgetary technical cooperation (XBTC) resources aimed at supporting outcomes. This complements the regular budget (RB) and RBSA, which can already be programmed flexibly towards DWCPs. This has allowed for the development of rigorous outcome-based programmes which can ensure project quality and consistency across regions. The next step will be to coordinate outcome level actions into coherent country packages in support of integral DWCPs. To do this, the ILO will strengthen the consultation mechanisms between outcome coordinators and field offices to see how outcome-based programmes can also work in synergy at the country level.	PARDEV/none	Three outcome-based partnership agreements were signed with Sweden (Phase II, 2012–13), Norway (2012–15) and Irish Aid (2012–15). Within this framework, PARDEV facilitates coordination between technical units, outcome-based coordinators, PROGRAM, EVAL and field offices with the aim of improving the design of project documents as well as results reporting based on decent work outcomes and CPOs.
		■ The tentatively scheduled high-level evaluation of the field structure review in 2013 will allow for an assessment of progress made in aligning country office structures and resources to support the DWCP approach.	EVAL/none	

Appendix II

Results-based management matrices for evaluation strategy

Outcome 1: Improved use of evaluation by ILO constituents and management for governance

Indicator		Baseline	End target
1.1.	The frequency and quality of EAC decisions and advice on relevance of evaluation programme of work to Governing Body policy decisions and strategic objectives of the Office; adequacy of follow-up to evaluation results	Three meetings in 2010; topics discussed for coming year only; no discussion of strategic use of evaluation recommendations	EAC convenes meetings and forums where analysis and dialogue on evaluation topics and follow-up lead to documented plans and follow-up for strategic use
1.2.	Annual evaluation report synthesizes recommendations and lessons learned based on evaluations	Reporting on implementation of evaluation strategy without analysis of broader ILO effectiveness	Annual evaluation reporting based on analysis of evaluation reports
1.3.	High-level evaluations assess the contributions of technical and decent work country strategies to the SPF and programme and budget outcomes	External quality rating of evaluations; 2005–09 (from independent external evaluation)	High-level evaluations better inform governance-level strategic and programming decisions

Biennial milestones

2010–11		2012–13	2014–15
1.1.	2011: EAC schedule, procedures and deliverables specified in new action plan; formal record of recommendations for evaluation programme of work (2012–13); record of EAC advice on use of specific recommendations	Four meetings per year; record of recommendations for evaluation programme of work (2013–14); record of EAC advice on use of specific recommendations	Four meetings per year; formal record of recommendations for evaluation programme of work (2015–16); record of EAC advice on use of specific recommendations
1.2.	Performance information in annual evaluation report based on analysis of evaluation reports; results discussed by Programme, Financial and Administrative Committee (PFAC)	2013: Improved annual evaluation report based on Governing Body feedback; results feed into the Programme and Budget for 2014–15	2015: Annual evaluation report used in developing new SPF and programme budget
1.3.	Results of internal peer review of high-level evaluations 2010–11 register satisfactory quality	Results of internal peer review of high-level evaluations 2012−13 register satisfactory quality	Results of external evaluation show high satisfaction with RBM link and usability of high-level evaluations 2010–15

Outcome 2: Harmonized Office-wide evaluation practice to support transparency and accountability

Indicator		Baseline	Target 2010–15
2.1.	By 2015, 100 per cent of DWCPs and projects would have mechanisms in place for regularly engaging constituents in the use of evaluation processes	n.a.	Results of periodic ex post surveys; reporting of constituent response and follow-up show 80 per cent of evaluations used by constituents; 100 per cent of final project reports document constituents' involvement and sustainability plans
2.2.	Upgrade and expand the use of evaluations for management (decentralized)	Count of self-, internal, thematic and impact evaluations conducted by sectors and regions (2009)	All regions and sectors have biennial evaluation plans that link to management accountability and organizational learning

Biennial milestones			
2010–11		**2012–13**	**2014–15**
2.1.	2011: Initial survey to constituents based on 2010 evaluations completed sets baseline measure	2013: At least a 25 per cent improvement in reported use of evaluations by constituents over 2011 levels	2015: At least a 50 per cent improvement in reported use of evaluations by constituents over 2011 levels
2.2.	2011: 20 per cent increase in use of self-evaluation to address organizational issues; 20 per cent use of project final progress report	All internal and self-evaluations accessible and searchable in the ILO's database	80 per cent use of project final progress report (self-evaluation) for projects above US$500,000; results of validation exercise measure validity and reliability of evaluation and reporting

Outcome 3: Evaluation capability expanded through enhanced knowledge, skills and tools

Indicator		Baseline	Target 2010–15
3.1.	Evaluation capacity and practice among ILO staff and constituents improved	Number of staff and constituents receiving technical training and hands-on support (2009)	By end of 2015, 225 additional constituents and 225 ILO officials develop specialized evaluation skills related to evaluation
3.2.	For evaluation network, standardized roles and responsibilities applied throughout the ILO	No standardized job descriptions for evaluation officers; compliance with evaluation guidelines unknown	Evaluation responsibilities specified in job descriptions; individual performance appraisals; roles and responsibilities standardized

Biennial milestones			
2010–11		**2012–13**	**2014–15**
3.1.	75 constituents and 75 ILO officials develop specialized knowledge through ILO training	75 constituents and 75 ILO officials develop specialized knowledge through ILO training	75 constituents and 75 ILO officials develop specialized knowledge through ILO training
3.2.	2011: ILO generic job descriptions developed for evaluation officers	2013: Internal governance document on evaluation network: approach, roles and responsibilities adopted and applied	