



CGIAR STANDARDS FOR INDEPENDENT EXTERNAL EVALUATION

“The [Evaluation] Policy is supported by a set of Evaluation Standards and a series of current Guidance Notes, issued by the Head – IEA, following full consultation with all pertinent stakeholders, in particular CRP management. These standards and guidance provide the details, modalities and common operating frameworks and standards for implementation of the Policy in the CGIAR.”

CGIAR Policy for Independent External Evaluation – paragraph 2

January 2015



Table of contents

List of Abbreviations.....	ii
Introduction	1
Purpose.....	1
CGIAR evaluations covered by these standards	2
Sources and derivation of the Standards	3
Evaluation principles	3
Evaluation Standards and the transition to a reformed CGIAR	4
Standards for planning and conduct of evaluations.....	5
Principle 1: Evaluation will be professional, conforming to internationally accepted standards and pursuing good practice	5
Principle 2: Quality Management will be applied to evaluation and facilitated through a Community of Practice	8
Principle 3: Evaluation will serve clearly defined target audiences.....	9
Principle 4: Evaluation will take account of the special characteristics of agricultural research for development in the CGIAR	9
Principle 5: Evaluation will help reinforce mutual accountability in the CGIAR system and with partners and beneficiaries.....	11
Principle 6: Managers in the CGIAR will reinforce evaluation relevance, follow- up, knowledge management and learning	12
Principle 7: Evaluation will be independent, ethical and transparent	14
Principle 8: Evaluation will be equity, gender and culture-sensitive	17
Principle 9: Evaluation will be efficient.....	17
Annex 1: Glossary of Evaluation Terms.....	19
Annex 2: Interpretation of International Evaluation Criteria	26

Tables

Table 1 Evaluations covered by the CGIAR Evaluation Policy and Standards.....	3
---	---

List of Abbreviations

CCEE	CRP-Commissioned External Evaluation
CRP	CGIAR Research Program
ECoP	Evaluation Community of Practice
IEA	Independent Evaluation Arrangement
ISPC	Independent Science and Partnership Council
REWP	Rolling Evaluation Work Plan
SPIA	Standing Panel on Impact Assessment
TOC	Theory of change
ToR	Terms of Reference

Introduction

1. The CGIAR Policy for Independent External Evaluation (“the Policy”) ¹ was approved by the Fund Council and the Consortium Board in 2012. It is part of the Common Operating Framework. The Policy sets out the mandate, scope and proposed implementation arrangements for independent external evaluation in the reformed CGIAR. These Standards are to be read in conjunction with the Policy, as they provide the details, modalities and a common operating framework for implementing the Evaluation Policy.
2. As stated in the Evaluation Policy (paragraph 2), the Policy “is supported by a set of Evaluation Standards and a series of current Guidance Notes, issued by the Head – IEA, following full consultation with all pertinent stakeholders, in particular CRP management. These standards and guidance provide the details, modalities and common operating frameworks and standards for implementation of the Policy in the CGIAR.” The Standards apply generally across all evaluations but much of the detail applies specifically to the evaluation of CGIAR Research Program (CRPs) or program components.
3. The Standards provide interpretation of and guidance for implementing the evaluation principles of the Policy. They contain two annexes aimed to promote consistency and harmonized approach to evaluations in the CGIAR. Annex 1 contains the Glossary of Evaluation terms, and Annex 2 presents an interpretation of the criteria applied in CGIAR evaluations and provides examples of evaluation questions specific to each criterion.
4. The Standards are supported by a set of Evaluation Guidance Notes which set out the suggested processes for conducting evaluation, approaches to specific types of evaluations and details on specific aspects and processes of evaluation.
5. The successful implementation of the Policy, however, depends on there being an evaluation culture (Principle 6 in the Policy) that is promoted by senior staff in governance and management across the CGIAR. Evaluations should be promoted as an opportunity for learning and improvement, and their results used as an important input to oversight and management.

Purpose

6. These Standards are a tool to promote external and independent evaluations in the CGIAR that are fit for purpose, i.e. useful, credible, accurate, ethical, and of high quality. They are intended primarily as reference by those planning, commissioning and carrying out evaluations, and form the basis for evaluation quality assurance. They also set out the

¹ The Evaluation Policy was adapted to the specifics of the CGIAR and drew from the Glossary of the OECD-Development Assistance Committee Evaluation Network and the Norms for Evaluation in the UN System, United Nations Evaluation Group, April.

<http://www.iea.cgiar.org/sites/default/files/CGIAR%20Evaluation%20Policy%20-%20Final%20approved%20document%20effective%20February%202012.pdf>

standards for contracting staff responsible for managing external evaluations in the CGIAR and evaluators conducting evaluations, and delineate the general responsibilities of those who manage evaluations, evaluators and those being evaluated.

CGIAR evaluations covered by these standards

7. The term ‘evaluation’ may be used for a wide variety of studies and reviews, but as defined in the Policy and with reference to OECD/DAC² it refers to “[..] the independent, systematic and objective assessment of an on-going or completed project, program, institution, policy or modality, its design, implementation and results. It determines the relevance and fulfillment of objectives, development efficiency, quality, effectiveness, impact and sustainability”. Evaluations are in-depth assessments that combine both formative and summative elements (for full definition, see Annex 1).

8. Table 1 below lists the main CGIAR evaluations covered by these Standards. These are primarily evaluations commissioned centrally, by the Head of the Independent Evaluation Arrangement (IEA), which is the body responsible for independent evaluation in the CGIAR. The Standards also apply to evaluations commissioned and managed by CRPs and Centers, as these are essential building blocks for the central CRP and CGIAR System-Wide evaluations. Each CRP is expected to commission a representative number of external evaluations (CRP-Commissioned External Evaluations - CCEEs) feeding into the CRP evaluation. For providing representative coverage of the CRP, the topics/scope of CCEEs are discussed and agreed biennially with the IEA Head (Policy: 4.3) and included in the overall CGIAR evaluation workplan (Policy: 6.1). CCEEs should conform to these Standards (Policy: 4.3).

9. There are many other evaluative studies in the CGIAR that do not fall within the oversight of the IEA, although they provide vital information on which IEA evaluations will draw. These include, for example: impact assessments commissioned by the Standing Panel on Impact Assessment of the CGIAR (SPIA) and within CRPs; most reviews and adoption studies commissioned within CRPs and Centers including by individual donors; Center Reviews commissioned by Center Boards or the Consortium Board; audits. These studies are not obliged to follow these Standards, but they could be of use to them since they reflect internationally-agreed principles and norms for evaluation.

² <http://www.oecd.org/development/evaluation/50584880.pdf>

Table 1 Evaluations covered by the CGIAR Evaluation Policy and Standards

Type	Managed by	How often
CGIAR System-Wide evaluation	IEA	Every 6-7 years
CRP evaluations	IEA	About every 5 years
CCEEs	CRP management / lead Center	Flexible: to feed into management decisions as well as overall CRP evaluation.
Evaluation of Central Institutions of the CGIAR System (Fund Council, Consortium, ISPC-SPIA)	IEA	Every 6-7 years, to feed into the System-Wide evaluation
Shared Services, including Gene Banks	Appropriate institutions of the CGIAR**.	Flexible: to feed into key decisions as well as System-Wide Evaluation.
Evaluation of the IEA	Externally (see Policy)	Every 6-7 years, to feed into the System-Wide evaluation
Demand-driven Evaluations on Specific Cross-CGIAR Issues and Themes	IEA	Flexible – expected less than one every two years

Source: Evaluation Policy 4.1-4.8.

Sources and derivation of the Standards

10. These Standards elaborate the principles in the CGIAR Evaluation Policy (see Introduction). They have been informed by common international evaluation standards and experience from research evaluation, particularly in the CGIAR^{3,4}. They have been developed in consultation with pertinent stakeholders, in particular the CRP management so that they are pragmatic in the current CGIAR context. Thus, the standards take into account the special characteristics of agricultural research and the experience of evaluation in the reformed CGIAR since 2013.

Evaluation principles

11. The basic principles for evaluation in the CGIAR are outlined in the approved Evaluation Policy (sections 3.2-3.10). These are:

- 1) evaluation will be professional, conforming to internationally accepted standards and pursuing good practice;

³ <http://www.oecd.org/development/evaluation/50584880.pdf>

⁴ http://www.sciencecouncil.cgiar.org/system/files_force/ISPC_ME_CGIAR.pdf?download=1

- 2) quality Management will be applied to evaluation and facilitated through a Community of Practice;
- 3) evaluation will serve clearly defined target audiences;
- 4) evaluation will take account of the special characteristics of agricultural research for development in the CGIAR;
- 5) evaluation will serve mutual accountability in the CGIAR system and with partners and beneficiaries;
- 6) managers in the CGIAR will reinforce evaluation relevance, follow-up, knowledge management and learning;
- 7) evaluation will be independent, ethical and transparent;
- 8) evaluation will be equity, gender and culture-sensitive;
- 9) evaluation will be efficient.

12. These Standards outline how the above principles should be interpreted in the planning, oversight and conduct of evaluation. Each standard is followed by some detail of interpretation, not all of which will be applicable to every type of evaluation. Quality assessments will focus on the spirit of the Standard, rather than whether a particular evaluation conforms to all the details set out below.

Evaluation Standards and the transition to a reformed CGIAR

13. The CGIAR is completing a transition to a reformed system, which includes many new institutions, including the IEA. Evaluations in the transition period will make use of processes and skills which are still under development. During this period, evaluations depend on information available from monitoring systems and reviews carried out prior to the reform, which may not be sufficient. Given this context, these Standards must be employed pragmatically, to encourage raising the quality and use of evaluations across the board, rather than with strict compliance.

14. During the transition, the CGIAR is expected to move to a more streamlined evaluation system. Firstly, the quality, coverage and utilization of CRP/Center evaluations will be assessed as part of preparation for the IEA-commissioned CRP evaluations and CGIAR System-Wide evaluation (Standard 2.2). Linkages between the evaluations that fall under the CGIAR Evaluation Policy and Standards, namely the IEA-commissioned ones, and CCEEs, are strengthened by work through the Evaluation Community of Practice (ECOP). The ECOP will help to support staff involved in CRP and Center evaluations, development of methodologies for evaluations and periodic updating of these Standards as deemed necessary (Standard 2.3). Secondly, the evaluation system will be strengthened by better use of evaluative studies conducted internally or commissioned by CRPs and Centers and studies conducted, for example, by SPIA. They can provide credible analysis of information

and primary data on which IEA and CRP evaluations draw - as well as being critical for learning and improvement within the research institutions themselves. The regular evaluation of the IEA (Standard 9.1) will provide an opportunity to assess the IEA role within the CGIAR System-wide evaluation.

Standards for planning and conduct of evaluations

Principle 1:

Evaluation will be professional, conforming to internationally accepted standards and pursuing good practice

Standard 1.1 Every evaluation will have clear Terms of Reference, covering objectives, scope, evaluation questions, approach and methodology, management and governance arrangements

15. The extent to which the Terms of Reference (ToR) can be tailored for individual evaluations depends on the type of evaluation. Because the CRP evaluations are similar in terms of their intervals, commissioning authority, users and types of stakeholders, general scope and evaluation criteria to be addressed, their ToR for CRP evaluations should be largely consistent to enhance comparability and harmonization. Also, CCEEs that are building blocks for CRP evaluations, should aim to harmonize ToRs. In other evaluations there is more flexibility. The generic coverage of ToR is given in separate guidance (Guidance Note 3)⁵.

Standard 1.2 Evaluation Inception Report will further elaborate on the Terms of Reference, particularly regarding the evaluation framework, approach and methodology

16. Evaluations will have an inception phase. During this phase, in collaboration with the IEA evaluation manager, the independent evaluation team leader will refine the evaluation framework and the methodology, develop methods, tools and a detailed plan for the evaluation. The team will review basic documentation and meet managers and other key stakeholders of the unit to be evaluated for gaining information and understanding, necessary for planning the evaluation. The inception phase will result in an Inception Report. The team can propose any changes they consider necessary to the ToR. Any changes agreed should be clearly documented and communicated, and they need to be agreed by the evaluation manager and specified in the Inception Report. The generic coverage for an Inception Report is given in separate guidance (Guidance Note 4)⁶.

⁵ <http://iea.cgiar.org/sites/default/files/G3.pdf>

⁶ <http://iea.cgiar.org/sites/default/files/G4.pdf>

Standard 1.3 Evaluation design will address the relevant international evaluation criteria

17. In the CGIAR, six key evaluation criteria are used: Relevance, Science Quality, Efficiency, Effectiveness, Impact and Sustainability. These correspond with international evaluation practice as established by OECD/DAC and applied in research context. Evaluation criteria serve as a check that all major issues are considered in an evaluation. The six criteria and their suggested interpretation and coverage are set out in the Annex 2 to these Standards, and should be consulted when developing evaluation questions.

18. In addition, there are criteria that are recommended to be used for institutional governance and management evaluations⁷. These are independence, accountability, transparency, legitimacy and fairness, in addition to efficiency and effectiveness.

19. An evaluation ToR may prioritize particular criteria, depending on the objectives of the evaluation and the nature and stage of the work being evaluated. The ToR should explain the reasoning behind the chosen priorities.

Standard 1.4 Selection and contracting of evaluators will follow international good practice

20. Evaluation teams consist of external evaluators. The team leaders, in particular, must have appropriate evaluation skills and experience, as well as sufficient understanding of research for development and expertise in the subject being evaluated. Normally, familiarity with the CGIAR will be a significant advantage, as long as conflict of interest is avoided (see Standard 7.2). Teams will also include appropriate skills in the fields of management and efficiency, and effectiveness of operations.

21. CGIAR evaluations require recruitment of external evaluators. Evaluators should be selected through a process that is internally transparent and documented, although information of individuals considered should not be published. The process should ensure that appropriately qualified assessors are involved in the selection, and that explicit criteria are used in selecting evaluators. In order to confirm the availability of highly qualified evaluators, sufficient time should be allowed for their identification, recruitment and contractual arrangements.

22. Evaluators' contracts are usually linked to the evaluation ToR. They should specify the component of the evaluation that they will primarily contribute to, to whom they will directly report, the process for approval of their contribution to the Evaluation Report, and the procedures for making any changes to the contract.

23. It is good practice that evaluators sign a declaration of interest, so that actual conflict of interest can be avoided and any potential or perceived conflicts of interest can be

7

<http://web.worldbank.org/WBSITE/EXTERNAL/EXTOED/EXTGLOREGPARPROG/0,,contentMDK:21178261~menuPK:4426473~pagePK:64829573~piPK:64829550~theSitePK:4426313,00.html>

managed. It is also good practice that evaluators sign a code of conduct, which covers issues of ethical and professional behavior in evaluation, such as intellectual property related to the evaluands and evaluation products, confidentiality required, and reporting in case evaluators uncover wrong-doing, fraud, or other misconduct⁸.

Standard 1.5 Evaluation methods will follow international professional standards

24. Using appropriate evaluation methods means:
- a) employing data collection and analytic methods (both qualitative and quantitative) which reduce - to the extent possible - subjectivity in judgments, and ensure that findings are based on systematic inference from evidential information and recommendations are documented by the evidence;
 - b) using appropriate methods, indicators and qualified individuals to judge specialized areas (see Standard 4.2);
 - c) triangulating information, assessing the reliability of secondary data and cross-checking personal opinions;
 - d) taking into account what would have been likely to have happened in the absence of the research or other activity being evaluated;
 - e) validating the quality of evaluative studies to be used as source of prior assessment and conclusions, such as impact studies;
 - f) taking into account different stakeholder perspectives (see Standard 5.1) including gender, cultural and generational perspectives;
 - g) taking into account risk and unpredictability which is an inherent characteristic of research and its effectiveness in development (see Standard 4.1);
 - h) taking the wider context into account. This means consideration of opportunity costs of the investment, the comparative advantage of those undertaking the work, and the appropriate degree of collaboration and coordination with others. It may also require benchmarking against similar programs;
 - i) systematic recording and maintenance of records from the evaluation, with appropriate care for confidentiality and protection of sources;
 - j) being explicit and transparent about strengths and limitations of the methodology, both during the evaluation and in Evaluation Reports.

Standard 1.6 Reporting of evaluations will follow international standards

25. The Evaluation Report should be logically structured, with a view to maximum clarity and interest for the target audience. It should contain findings, conclusions, lessons and

⁸ IEA declaration of interest and code of conduct; <http://iea.cgiar.org/sites/default/files/interest%2Bcode.pdf>

recommendations with clear links between them and showing how they derived from the analysis of evidence and information⁹.

26. Inputs, activities, outputs, outcomes and impacts should be clearly documented wherever relevant to the scope of the evaluation. Sources of data and information should be referenced. Institutional and organizational set-up and responsibilities and processes at each level should be clearly presented. The generic coverage for an Evaluation Report is given in separate guidance (Guidance Note 5)¹⁰.

Principle 2:

Quality Management will be applied to evaluation and facilitated through a Community of Practice

Standard 2.1 The quality of all evaluations commissioned directly by the IEA will be subject to assessment according to these standards

27. The IEA is responsible for quality assurance of individual evaluations that it commissions. The quality assurance across evaluations should be consistent.

28. The IEA senior staff assures the quality of TOR of the IEA commissioned evaluations. Evaluation managers have a responsibility for ensuring high-quality evaluation, and quality assurance is one of their roles.

29. The IEA calls on evaluation peers for quality assurance advice on various aspects of an individual evaluation, in particular on the draft Inception Report and draft Evaluation Report. IEA may also use Expert Panels to comment on the quality and feasibility of findings, conclusions and recommendations of Evaluation Reports. However, the ultimate responsibility for evaluation quality rests with the Head of the IEA.

Standard 2.2 The quality of CCEEs, and any other decentralized evaluations, will be assessed in the framework of CRP and system-wide evaluations.

30. The CGIAR's Evaluation Policy introduces a hierarchy of evaluations, in which CCEEs serve as essential building blocks for CRP evaluations. These, along with evaluations of CGIARs governing and system units, serve as inputs into system-wide evaluations.

31. The quality of evaluations being used as inputs into "higher-level" evaluations will be assessed during the preparatory phase of the CRP or system-wide evaluation.

32. With regard to CCEEs, consultation between the IEA and the CRP management and evaluation staff during planning and implementation of CCEEs also provides an opportunity for advising CRPs on aspects of evaluation quality.

⁹ See Guidance notes G1 (<http://iea.cgiar.org/sites/default/files/G1.pdf>) and G5 (<http://iea.cgiar.org/sites/default/files/G5.pdf>)

¹⁰ <http://iea.cgiar.org/sites/default/files/G5.pdf>

Standard 2.3 External Independent Evaluators and the Evaluation Community of Practice will have an input into updating these Evaluation Standards

33. The Head of IEA will ensure that these Standards are periodically assessed regarding their appropriateness and application to evaluation quality management and updated as deemed necessary. This should happen after completion of a cycle of evaluations including the System-wide evaluation, thus every 6 years. This assessment will explicitly seek the views and experiences of independent external evaluators and the EcoP (see Evaluation Policy 3.3). The latter is an important stakeholder group to the IEA to assess the applicability of the Standards, implement evaluations following these Standards, and consider revisions to the Standards deemed necessary overtime.

Principle 3:

Evaluation will serve clearly defined target audiences

Standard 3.1 The target audiences and expected timing and use of the evaluation will be taken into account in the planning and design of evaluations

34. Evaluations should be timed so that they can contribute to decision-making wherever possible.

35. ToRs should clearly define the target audiences and expected use of the evaluation in order to be relevant and useful for both decision-making and learning. In the design of the evaluation and development of ToR, the requirements of decision-makers will be taken into account.

Principle 4:

Evaluation will take account of the special characteristics of agricultural research for development in the CGIAR

Standard 4.1 The unpredictable and risky nature of research, and the long time it takes for research to result in development outcomes will be taken into account in the planning and conduct of evaluations

36. In the evaluation planning and approach, the special characteristics of research, and research for development in particular, will be taken into account. These include unpredictability and risk related to research, need for innovation and ambition, the protracted impact pathways from research to outcomes, and serendipity related to outcomes. Furthermore, outcomes and impact depend on factors that are not in the control of the research programs.

37. Evaluation of research programs and program components will check progress against plans underlined by research impact pathways, and the extent to which research hypothesis and results are used for adjusting plans. Thereby evaluation will support learning and adaptive management.
38. Evaluation planning needs to take into account the type and stage of research being evaluated.
39. Evaluations should, wherever relevant:
 - a) assess the validity of the assumptions underlying ‘impact pathways’ and theories of change of the research in terms of the evidence base or there being appropriate hypotheses to be tested, and to assess that the entity being evaluated has adequate systems and procedures to test those assumptions and to generate new evidence. Depending on the nature of research, the evidence base underpinning the theories of change may be limited, and the focus of the evaluation will be on the knowledge and evidence to be generated through research;
 - b) ask questions about how research that has not produced results as expected or has had different results than anticipated is documented and used in learning. This is to incentivise researchers to present and learn from the full range of results, from successes to failures, and to prevent skewing the overall conclusions from systematic reviews of research;
 - c) include questions about the processes for risk management in research.

Standard 4.2 The specialist and development-focused nature of CGIAR research will be reflected in the selection of evaluators and evaluation methods

40. Expertise on research processes and research management and on the specific subject areas being evaluated must be adequately represented in the evaluation team. The team leader must have a strong understanding of research for development, as well as solid evaluation skills and experience. Scientific panels and/or peer reviewers may be used to provide additional specialist expertise. If a separate panel is used, clear arrangements need to be set out on how its input will be integrated with the rest of the team’s work.
41. Evaluations tools and methods should be chosen so as to assess results performance against appropriate management of research ambition, risk and failure.
42. Evaluation of research quality should be based on a minimum set of criteria (see Annex 2), and draw from the subject matter expertise of the team members and additional scientific panels, when these latter are employed.
43. Measurements and indicators, both qualitative and quantitative, may be available in the CRP and CGIAR monitoring systems. When such indicators are used, the evaluators will need to verify their validity and robustness as measures for judging program performance,

or benchmarking a program against others, and consider explicitly the strengths and limitations of any indicators used.

44. The IEA should promote development and use of harmonized indicators across evaluations and for evaluating specific aspects, for instance science quality in CRPs.

45. Evaluations should assess the use of indicators in program decision-making, including whether performance monitoring has appropriate incentives on research.

Standard 4.3 The objectives of the CGIAR and of the CGIAR reforms will be reflected in evaluation design

46. To the extent possible, evaluations should examine the effect of the evaluated program, institution or activity in contributing to the overall objectives of the CGIAR following its reform. Evaluations will also take into consideration the comparative advantage the program has compared to other research suppliers.

47. Regarding non-research bodies of the CGIAR, evaluations should consider the comparative advantage they have in delivering their planned outputs.

48. Evaluations should maintain an independent stance towards the implementation of the CGIAR reform, and assess the effects both intended and unintended of the reform process on the planned objectives programs evaluated (or successful functioning of other units/functions evaluated).

Principle 5:

Evaluation will help reinforce mutual accountability in the CGIAR system and with partners and beneficiaries

Standard 5.1 Planning and conduct of evaluations will take into account the role of key stakeholders

49. The managers of programs being evaluated should assist evaluation planners with the identification of a realistic number of 'key stakeholders', and in outlining their potential roles and responsibilities in evaluation.

50. For individual evaluations, a small group of key stakeholders should be consulted on major design questions, for instance evaluation rationale, evaluation questions and intended use of the evaluation results. This will be through participation in a reference group if such a group is established¹¹. Key stakeholders should also be given adequate opportunity to provide input into the evaluation work program as a whole.

51. Evaluation ToR should be shared and/or communicated in an appropriate format to key stakeholders.

¹¹ See Guidance Note 4 on CRP evaluations: <http://iea.cgiar.org/sites/default/files/G4.pdf>

52. Consideration must be given to how best to elicit stakeholder inputs during the evaluation process, and if necessary how to encourage groups to organize themselves. Interim evaluation findings should be disseminated to key stakeholders. Their comments and feedback should be taken into consideration in the production of the Evaluation Report.

Standard 5.2 The governance arrangements for evaluations will make role and responsibilities of different stakeholders clear

53. The role and responsibilities of different stakeholder should be clarified and recorded, for example in the evaluation ToR, and in the Inception Report¹².

Standard 5.3 Evaluations will investigate the performance of key partnerships and different parties in the partnership

54. In the spirit of mutual accountability, each entity within the CGIAR system is accountable to the others and, as defined in the CGIAR Strategic Results Framework, accountable to the ultimate beneficiaries. Evaluations will therefore look at the performance of different parties, including funders, in CGIAR partnerships to the extent feasible. Evaluation recommendations may be directed at funders and other partners as well as researchers and other implementers.

55. Managing evaluations jointly with one or more key partners can increase their legitimacy and promote learning and capacity building, as well as serve the purpose of the CGIAR reform to establish a single evaluation system. Norms for such jointly-managed evaluations are listed in separate guidance.

Standard 5.4 The CGIAR evaluation system addresses accountability at different levels of CGIAR operations

56. The evaluation system, which covers both programs and governing bodies and units of the CGIAR and the overall CGIAR system, will assess accountability mechanisms at different levels in the system, and for both oversight and management of research. The processes that are part of mutual accountability include setting priorities in consultation with the partners and beneficiaries at CGIAR and CRP level, adequately funding of prioritized research, appropriate allocation and expenditure of the funds for relevant and high quality research and research management, monitoring of progress and adaptive management, and documentation of outputs, outcomes and impacts.

Principle 6:

Managers in the CGIAR will reinforce evaluation relevance, follow-up, knowledge management and learning

¹² See Guidance notes 3 (<http://iea.cgiar.org/sites/default/files/G3.pdf>) and 4 (<http://iea.cgiar.org/sites/default/files/G4.pdf>)

Standard 6.1 The CGIAR will promote a 'culture of evaluation' that actively plans for and welcomes evaluations, and promotes constructive learning

57. Senior staff in governance and management across the CGIAR (including the Consortium and Fund Council) has the important responsibility to create a culture and constructive environment for evaluations. This includes: promoting high quality monitoring and evaluation systems that form the building block of higher order evaluations, promoting evaluations as an opportunity for learning and improvement, rather than for penalising 'failures', receiving evaluation results as an important input to oversight and management and ensuring that sufficient resources are allocated to work with evaluators, so that evaluation does not constitute an unacknowledged additional burden on staff. Evaluations should be seen as an opportunity to foster dialogue within the program/unit and to collectively identify lessons.

Standard 6.2 Evaluations will be communicated publicly and in appropriate formats for stakeholders

58. The evaluation results should be presented in accessible formats to key stakeholders and systematically distributed internally and externally. This should include oral presentation through meetings and the media as appropriate. Communications to specific stakeholders should include all important findings, conclusions and recommendations that bear on the interests of those stakeholders. Publishing evaluation briefs in the same language as that of important stakeholder groups should be done. Presentation should allow for feedback loops from stakeholders and their representatives to facilitate the follow-up on evaluation results and implementation of the recommendations.

59. All final CGIAR Evaluation Reports falling under the Policy (see Table 1) will be made publicly available via the internet on the CGIAR-IEA website within three months of their completion. Other evaluative studies will also be made public where feasible, including any evaluations managed by CGIAR partners.

60. All essential elements of IEA-managed evaluations will be publicly available on the internet, including: the IEA Rolling Evaluation Workplan and Budget (REWP); evaluation Terms of Reference; Inception Reports; Evaluation Reports; management responses and follow-up reports.

Standard 6.3 There will be a formal and systematic management response to each evaluation

61. The relevant managers must respond to the evaluation and its specific recommendations within a reasonable timeframe. The management response will be made publicly available.
62. The governing body for the program or institution being evaluated is responsible for consideration of the evaluation findings and recommendations, together with the management response to these, and for approving the evaluation follow-up action plan.
63. Progress on agreed follow-up actions will be reported on by management. Evaluations of CRPs and of the CGIAR as a whole will also cover the extent to which agreed actions have been followed up.

Principle 7:

Evaluation will be independent, ethical and transparent

Standard 7.1 Appropriate safeguards will be in place for CGIAR evaluation staff to protect and promote structural and behavioural independence

64. Staff in CRPs and Centers who has responsibilities for evaluation should report directly to the body overseeing these functions, in addition to reporting to the immediate supervisors. This staff should have adequate orientation and training on code of conduct of evaluations.
65. CGIAR evaluation staff should not influence the findings, analysis or recommendations of Evaluation Reports written by independent evaluators. In their role as providing quality assurance to evaluations, evaluation staff is responsible for overseeing that the Evaluation Report meets IEA quality standards, while, in doing so, respecting the independence of the evaluation team.

Standard 7.2 The evaluation team will be independent of the work being evaluated, with any potential or perceived conflicts of interest being transparently managed

66. In order to avoid any real or perceived bias due to conflict of interest or other factors (such as personal competition among researchers, or a tendency to advocate a particular specialisation), the following points are important:
 - a) independence and impartiality are essential criteria in selecting evaluation teams;
 - b) people on evaluation teams should not be evaluating their own work in any way. The following constitute a conflict of interest that prevents people from serving in an evaluation team: engagement during a defined period (5-10 years) in a given program, unit or work being evaluated as staff member, consultant, member of

- oversight or advisory body, donor, implementing partner or in any other role that would render in any way the contribution made subject to the evaluation;
- c) evaluation managers should canvas broadly, for example using contacts to existing science panels and evaluation groups for establishing a pool of potential evaluators;
 - d) where possible evaluation teams should represent a diversity of backgrounds (see Standard 8.1);
 - e) candidates for evaluation teams must declare in writing any potential conflict of interest (real or likely to be perceived) and the evaluation manager will consider case by case whether the cause of potential conflict of interest prevents the candidate from being selected. Actual conflict of interest discovered during the evaluation should lead to termination of the contract;
 - f) any potential conflict of interest which arises during the evaluation should be dealt with and managed transparently involving program management if needed, so that it does not compromise the evaluation processes, results or management response to the evaluation.

Standard 7.3 The conduct of evaluations will follow international ethical principles

67. Evaluators should be made aware of, and agree on, a set of core principles that constitutes the code of conduct for the evaluation¹³. Ethical principles for the conduct of evaluations include:

- a) evaluators will follow the principle of informed consent and respect people's right not to engage, to withdraw at any time and not to answer specific questions;
- b) evaluators should respect people's right to provide information in confidence, and must ensure that sensitive information cannot be traced to its source. If evaluators wish to quote someone directly, they should ask for permission, and people should be given a chance to review any quotations attributed to them;
- c) interviews should be conducted to minimize potential biases both of evaluators and informants. Evaluators should aim to create a safe and open atmosphere, where confidentiality is respected;
- d) evaluators should avoid offending the dignity and self-respect of those persons with whom they come in contact in the course of the evaluation;
- e) evaluators should provide maximum notice and minimize their demands on the time of interviewees (including researchers). They should avoid scheduling interviews in places and at times which will make it more difficult for certain groups (for example women) to participate;

¹³ The IEA has adapted the Code of Conduct for Evaluation in the United Nations System by the United Nations Evaluation Group. <http://iea.cgiar.org/sites/default/files/interest%2Bcode.pdf>

- f) knowing that evaluation might negatively affect the interests of some stakeholders, evaluators should conduct the evaluation and communicate its purpose and results in a way that clearly respects the stakeholders' dignity and self-worth;
- g) evaluators should acknowledge significant contributions to the evaluation, unless anonymity is requested, including appropriate reference to any published or unpublished documents;
- h) if client requests conflict with ethical and professional standards of evaluators, evaluators should explicitly identify and discuss the conflicts with the client and relevant stakeholders, resolve them when possible, determine whether continued work on the evaluation is advisable if the conflicts cannot be resolved, and make clear any significant limitations on the evaluation that might result if the conflict is not resolved.

Standard 7.4 Evaluations will specifically consider any ethical issues relating to the program under evaluation.

68. These include, for example:

- a) honesty and transparency, including in reporting negative results and 'failure';
- b) research with human subjects, for example in health and nutritional studies;
- c) research with animals;
- d) benefits from and access to intellectual property including publications, patents and genetic material and sharing of benefits with national and local stakeholders;
- e) use and treatment of indigenous and traditional knowledge;
- f) ethical issues relating to certain types of technology or research, for example gene technology and randomized control trials (RCTs);
- g) considerations of equity, gender, generational differences, scale differences, conflicting interests etc. in the outcomes of the research.

69. Where relevant, evaluators should also consider what internal processes exist to promote an ethical approach in the program being evaluated (for example the existence of an ethics policy and ethics committee).

Standard 7.5 Evaluations should not be used for decision-making in individual human resource or audit matters

70. Evaluators are not expected to evaluate individuals, and must balance an evaluation of management functions with this general principle. If evaluators inadvertently uncover misconduct or fraud in the course of their work, this should be reported to the appropriate body.

Standard 7.6 Differences of opinion will be handled transparently

71. Relevant stakeholders must be given the opportunity to comment on the draft report and recommendations. Stakeholders', including management, comments should focus on factual errors and inconsistencies, which the evaluation team should rectify and reassess to the extent possible. The independent evaluation team has the final say on the final report wording. Evaluation team members should have the opportunity to dissociate themselves from particular judgments and recommendations on which they disagree. Any unresolved differences of opinion within the team should be acknowledged in the report.

Principle 8:

Evaluation will be equity, gender and culture-sensitive

Standard 8.1 Evaluation teams will insofar as possible be gender balanced, geographically diverse and include professionals from the countries or regions concerned

72. It is desirable that evaluation teams reflect gender and geographical balance. However, it will not always be possible to achieve sufficient diversity in the evaluation team, given the small size of the evaluation teams, and the need for specialist skills and evaluation competence. In this case, evaluators must work with evaluation managers to find appropriate mechanisms to adequately reflect diverse viewpoints.

Standard 8.2 Evaluators should address different stakeholder perspectives, including gender, cultural and generational perspectives

73. Evaluators must be sensitive to and address any issues of discrimination and inequality.

74. Evaluations should use disaggregated data to aim to capture inputs, processes and outcomes that differ by gender or social group. Evaluations covered by these standards are only rarely collecting primary data, so they will need to assess the extent to which disaggregated data exist, and use such data, as well as assessing to what extent research planning, monitoring systems and CRP-level evaluative studies address the research needs and potential benefits of different social groups.

Principle 9: Evaluation will be efficient

Standard 9.1 The IEA evaluation program as a whole will be balanced, representative and efficient

75. The IEA REWP includes the IEA commissioned evaluations (see Table 1). The process of developing and agreeing the REWP should promote a realistic use of available financial and human resources. The REWP is discussed with a range of key stakeholders (in particular

Centers, CRPs and funders), and the workplan and budget is then approved biennially by the Fund Council after receiving comments from the Consortium. The detailed procedure for developing and approving the REWP is set out in separate guidance.

76. The external independent evaluation of the IEA will examine the effectiveness and efficiency of the IEA and the usefulness of the evaluations and processes that it commissions and promotes. The evaluation will consider the IEA's work in the context of the entire evaluation effort of the CGIAR and the relative resource needs and benefits of different types of monitoring and evaluation activity, including learning and other evaluative activities conducted by CRPs and Centers.

Standard 9.2 Evaluations will be cost-efficient

77. Costs accrued from evaluation to CRPs and other units should be minimized, and value added to all stakeholder groups maximized. This can be achieved, for example, by: good advanced communication of forthcoming evaluations, whom they concern and in which ways; centralized collection of data and information with clear purpose as to why they are collected; effective feed-back on evaluation and communication of results to all stakeholder groups concerned, for example through briefs; reduction of parallel donor reviews through dialogue and communication about CGIAR evaluations.

Standard 9.2 The scope and outputs of evaluations will be realistically specified and achievable within the limitations of time, information and budgets

78. For each evaluation, the ToR - and the Inception Report in more detail - should set out clear and realistic expectations of what the evaluation will and will not cover.

Standard 9.3 Evaluation recommendations will be focused and practical

79. Evaluation recommendations should be action-oriented, practical and specific. There is no set limit on numbers of recommendations, but they should be focused on priority issues for actions by management or governing bodies. More detailed working-level suggestions may be made by the evaluators, for instance in annexes, but will not have the status of recommendations with an official response and follow-up (see Standard 6.3).

Annex 1: Glossary of Evaluation Terms

The purpose of this glossary is to promote consistent use in the CGIAR of terms and definitions related to evaluation. It draws from the OECD-DAC *Glossary of Key Terms in Evaluation and Results Based Management*¹⁴ and the World Bank *Sourcebook for Evaluating Global and Regional Partnership Programs*¹⁵ among other sources. The Independent Science and Partnership Council (ISPC) white paper *Strengthening Strategy and Results Framework through Prioritization*¹⁶ has also been consulted. The definitions take into account the use of the terms in the context of evaluating research for development and the CGIAR, in particular.

Accountability: Accountability is a concept used with many definitions and in many contexts (bureaucratic, political, professional, legal). In governance and management it can be defined as the duty to ensure and report that the use of authority is aligned with rules, standards, policy and interests of the program, organization and the broader group of stakeholders. Its assessment looks at the extent to which it is defined, accepted, and exercised along the chain of command and control within a program. In the context of research for development, accountability can be seen the obligation to take responsibility for performance in light of commitments, to the extent that performance is in the control of the program/project. Accountability requires ownership and acceptance of responsibility and the ability to deliver or influence the delivery of the desired results.

Mutual accountability: In the context of the CGIAR, this refers to the accountability of all partners, including donors, for the efficiency of outputs, outcomes and impacts of a program, institution or policy and sustainability of research.

Adoption, use, uptake: In research for development impact pathways, adoption, use and uptake refer to the primary and direct use of the research output by beneficiaries. This use may lead to changes; i.e. outcomes and impacts. Influence is a related term that implies an effect of research results on the primary beneficiaries or course of events.

Appraisal: An *ex-ante* assessment of the quality, relevance, feasibility and potential for impact and sustainability of a research program or activity, usually prior to a decision on funding it.

Attribution: The causality between observed (or expected) changes and an output from research or related activity. Attribution refers to both isolating and estimating the particular contribution of a program/project to the outcome/impact.

Audit: Financial and management audit in the CGIAR provides accountability to management at the level of the Center Boards, Consortium and Fund Council on finances

¹⁴ <http://www.oecd.org/development/peer-reviews/2754804.pdf>

¹⁵ <http://siteresources.worldbank.org/EXTGLOREGPARPROG/Resources/sourcebook.pdf>

¹⁶ http://ispc.cgiar.org/system/files_force/ISPC_WhitePaper_Prioritization.pdf?download=1

and assets and also provide elements of oversight in human resources and business efficiency.

Baseline: An analytical description of the situation prior to research activities, against which progress can be assessed or comparisons made.

Behavioral independence: see Independence

Beneficiaries: The individuals, groups, or organizations, whether targeted or not, that benefit, directly or indirectly, from the chain of events that research has contributed to (see also Target Group).

Clients: In these standards this term refers to the primary clients of evaluation – those requesting or receiving the evaluation (for example senior managers or donors of a CRP). Elsewhere it is often used to refer to the target group for a research project.

Comparative advantage: In economic terms, a comparative advantage in producing or selling a good is possessed by an individual, firm or country with the lowest opportunity cost (as opposed to absolute cost) in producing the good. In these standards the term refers more broadly to the role and mandate of the CGIAR in producing international public goods where there are no alternative research suppliers that are better positioned to produce those goods.

Cost effectiveness: The extent to which the program has achieved or is expected to achieve its results at a lower cost compared with alternatives. Cost-effectiveness analysis is distinct from cost-benefit analysis, which assigns a monetary value to the measure of effect. In research programs costing of outputs is more feasible than outcomes that typically depend on conditions and activities outside of research.

Counterfactual: The situation or condition which (hypothetically) would have prevailed if there had been no program/project.

Effectiveness: The extent to which the program or project objectives were achieved, or are expected to be achieved, taking into account the exploratory nature and risks inherent to research.

Efficiency: A measure of how economically resources/inputs (funds, expertise, time, etc.) are converted to results. In research context, assessment of efficiency refers to activities and outputs that are in the control of the research programs or cut across several CRPs and takes into account the exploratory nature and risks inherent to research. In the private sector “value for money” is commonly used for efficiency.

Evidence: The information presented to support a finding or conclusion. Such evidence should be sufficient and relevant. There are several sources for evidence: observations

(obtained through direct observation of people or events); documentary (obtained from written information); analytical (based on computations and comparisons); self-reported (obtained through, for example, surveys) and experiential (based on professional understanding and expertise that is accumulated over time) and is based on credible and legitimate sources.

Evaluation: The systematic and objective assessment of an on-going or completed project, program or policy, its design, implementation and results. In the CGIAR evaluation refers to an external, completely (IEA commissioned) or largely (CRP commissioned) independent and systematic study of an in-depth nature that uses clear evaluation criteria. In addition to research, it applies also to central CGIAR institutions, support programs and themes, and the System as a whole. An evaluation should provide information that is credible and useful, enabling the incorporation of lessons learned into the decision-making processes of major stakeholders.

Formative evaluation focuses on program/project implementation and is improvement-oriented.

Summative evaluation focuses on assessing worth of the program/project and lessons learnt (results and consequences), for instance to enable assessments with respect to change, continuation or enlargement of the program/project.

Evaluation criteria: Different aspects of quality of a program which are used internationally to develop evaluation questions and serve as a check that all major issues have been considered. In the CGIAR these include relevance, efficiency, quality of science, effectiveness, impact and sustainability.

Evaluation reference group: A structure set up to work with the evaluation managers to ensure good communication with, learning by, and appropriate accountability to primary evaluation clients and key stakeholders, while preserving the independence of evaluators.

Global public goods: These are defined as goods with the three following economic properties: 'non-rivalrous' (i.e. consumption of this good by anyone does not reduce the quantity available to others), 'non-excludable' (it is impossible to prevent anyone from consuming it) and available worldwide. In the CGIAR also the term International public goods is used. It refers to issues that are deemed to be important to the international community; and typically cannot, or will not, be adequately addressed by individual countries or entities acting alone.

Impacts: Positive and negative, primary and secondary long-term effects resulting from a chain of events to which research has contributed, directly or indirectly, intended or unintended. These effects can be economic, socio-cultural, institutional, environmental, technological or of other types. Note that sometimes the term impact is used to refer to more immediate results, here defined as Outcomes.

Impact assessment: In the CGIAR this term is generally used for an *ex post* study which uses specialized methods to estimate the changes in selected development parameters and the extent to which these are attributable to defined research activities or programs of the CGIAR. The Standing Panel on Impact Assessment (SPIA) has an oversight and capacity building function for impact assessment in the CGIAR.

Impact evaluation: The systematic assessment of the effects - positive or negative, intended or unintended - of program/project output(s) or intervention(s) on the outcomes/impacts of the affected groups and environments, and the extent to which these outcomes/impacts can be attributed to the program/project.

Impact pathway: The causal pathway for a research project or program that outlines the expected sequence to achieve desired objectives beginning with inputs, moving through activities and outputs, and culminating in outcomes and impacts. Assumptions underpinning the causal chain and feed-back loops are usually included (Closely related terms include Logical Framework and Theory of Change.)

Impartiality: In conducting an evaluation, the absence of bias in due process, in the scope and methodology, and in considering and presenting achievements and challenges. The principle applies to the clients of the evaluation, donors and partners, management, beneficiaries, and the evaluation team.

Independence: An evaluation that is carried out by entities and persons free from the control of those involved in policy making, management, or implementation of program activities. This entails both **organizational and behavioural independence**, protection from interference, and avoidance of conflicts of interest.

Organizational independence: Independence of the evaluation from management of the entity being evaluated. This encompasses functional independence (no conflict of interest), independent reporting line and communication right, guarantees against retaliations, independent budgeting and full access to information.

Behavioral independence: Objectivity and impartiality on the part of evaluators (which is not guaranteed by structural independence; for example evaluators may be reluctant to be critical of people they think may provide them with future contracts).

Indicator: A quantitative or qualitative variable that represents an approximation of the characteristic, phenomenon or change of interest (for instance, efficiency, quality or outcome). Indicators can be used to monitor research or to help assess for instance organizational or research performance.

Inputs: The financial, human, and material resources used in research.

Intermediate development outcome (IDO): At CGIAR's research program level targets representing CRP-specific thrusts and target domains that are generated as a result of

multiple activities by diverse actors outside the CGIAR. Their scales reflect CRP target domain and estimated volume of benefits. At System level IDOs represent accumulation of CRP outcome results with the scale corresponding to the CGIAR's target domains.

International public goods: – see Global public goods

Legitimacy: As a criterion for assessing governance and management, the way in which governmental and managerial authority is exercised in relation to those with a legitimate interest in the program — including shareholders, other stakeholders, implementers, beneficiaries, and the community at large.

Monitoring: A process of continuous or periodic collection and analysis of data to compare how well a project, program, or policy is being implemented against expected progress and results, in order to track performance against plans and targets, to identify reasons for under or over achievement, and to take necessary actions to improve performance. Monitoring is usually the responsibility of program management and operational staff, while evaluation as defined in the CGIAR Policy and Standards is carried out by external evaluators. Monitoring is also used for research purposes to guide decisions on research design and adjustment.

Organizational independence: See independence

Outcome: The intended or unintended short-term and medium-term effects resulting from an intervention's outputs. Note that the term impact is sometimes used as a generic term for effects, including the short-term results, here defined as Outcomes. See also Research outcomes.

Outputs: The products, new knowledge and services which result from research, capacity building and other activities related to research for development.

Peer review: A process of review involving qualified individuals within the relevant field. Peer review methods are employed to maintain standards of relevance and quality to improve performance and provide credibility. A peer review may be an input into an evaluation.

Performance management: The continuous process of setting goals, measuring progress, giving feedback, coaching for improved performance, and rewarding achievements

Performance measurement: The ongoing monitoring, measurement and reporting of program accomplishments and progress toward pre-established goals, which involved collecting data on the level and type of activities (inputs) and the products and services delivered by the program (outputs).

Relevance: The extent to which the objectives of a development intervention are consistent with global and national priorities and policies, as well as those of intended beneficiaries, partners and donors. In these Standards, it also refers to the extent to which the program is consistent with the goals, the System Level Outcomes, comparative advantage and reform agenda of the CGIAR, and program activities are consistent with the objectives of the program and its Intermediate Development Outcomes.

Research outcomes: The effects from research outputs applied by intermediary users, for instance by national partners or international research or development organizations.

Results: The output, outcome or impact (intended or unintended, positive and/or negative) of an intervention.

Results-based management: A management strategy focusing on performance and achievement of outputs, outcomes, and impacts.

Review: An assessment of the progress and performance of an intervention (including research), periodically or on an *ad hoc* basis. The words evaluation and review are often used interchangeably, but in the CGIAR, an evaluation refers to an external, completely (IEA commissioned) or largely (CRP commissioned) independent and systematic study of an in-depth nature using clear evaluation criteria, whereas reviews may be more flexible and narrow in focus.

Scaling up and scaling out: In agricultural development the terms are used nearly interchangeably to refer to the expansion of beneficial impacts from agricultural research and rural development. Scaling up/out relate to expanding, replicating, adapting, and sustaining successful policies, programs, or projects in geographic space or over time to reach a greater number of people. Scaling is typically preceded by piloting the model, idea or approach initially in a small scale. Scaling-out may refer specifically to the adoption and adaptation to local circumstances by users; while scaling-up may refer to extension and institutional support related to scaling.

Stakeholders: Agencies, organizations, groups or individuals who have a direct or indirect interest in the CGIAR or its component, for instance research program or its evaluation.

Sustainability: The continuation of benefits from a program intervention after research has been completed; the probability of continued long-term benefits or scalability of the benefits; the resilience to risk of the net benefit flows over time.

System Level Outcomes: The high level impact goals of the CGIAR: Reduction in rural poverty; Increase in food security; Improving nutrition and health; More sustainable management of natural resources.

Target group: The individuals or organizations for whose benefit the research or activity is ultimately undertaken, for example farmers or consumers in particular regions or agro-ecologies. (The word client can also be used, but is reserved in these Standards for primary clients of evaluation.)

Theory of change (TOC): Presents a hypothetical identification of the ways by which change is expected to occur from output to outcome and impact along an impact pathway. The TOC questions the assumptions about causality underlying the relationships between outputs, outcomes and impact. In TOC the assumptions present the mechanisms of change. There is no single method or presentational form agreed for TOCs. In research it is often used as a framework for testing hypotheses and incrementally building up the evidence base for the assumptions.

Transaction cost: The costs of planning, adapting and monitoring tasks completion. Transaction cost analysis includes comparison of transaction costs under alternative governance or operating structures.

Triangulation: The use of different sources or types of information, evaluators or types of analysis, to verify and substantiate an assessment, in order to overcome the potential bias that comes from a single source or method.

Annex 2: Interpretation of International Evaluation Criteria

Evaluation criteria are used internationally in order to encourage a systematic approach and coverage of all aspects considered important in evaluation to be assessed.

The evaluation criteria to be used in CGIAR evaluations include the agreed international evaluation criteria as per OECD-DAC. These criteria have been used in the CGIAR together with an additional criterion to reflect the research mandate of the CGIAR. As a result, in CGIAR evaluations the six main criteria are: **relevance, science quality, efficiency, effectiveness, impact and sustainability.**

The criteria are complementary and should be used together. Criteria may overlap in places and common sense should be used in addressing them. An evaluation may prioritize particular criteria, depending on the objectives of the study and the stage of the work being evaluated. As discussed in the Standards, additional criteria may be used, for example to assess governance and management (such as transparency and independence) that are not included here.

This annex sets out an interpretation of the main evaluation criteria for the CGIAR particularly as they apply to the evaluation of research programs¹⁷. In evaluation of the CGIAR's central institutions, for instance, the interpretation of the criteria and choice of evaluation questions will be different and the criteria for science quality and impact may not apply. For each criterion, specific evaluation questions need to be determined reflecting its interpretation, taking into account i) the similarities across programs and research, ii) requirement of consistency across evaluations, and iii) the specific nature of the program or work evaluated. Below, examples of evaluation questions are given.

Relevance

Relevance is the extent to which the objectives and design of the program or intervention being evaluated fit with (a) current global/regional policies, challenges and concerns (b) the needs, policies and priorities of intended beneficiary countries and groups; (c) the international and national research environment and research results being generated elsewhere (d) the specific objectives, role and comparative advantages of the CGIAR (including new opportunities arising from technology or partnerships); and (e) the specific objectives of the program of which a part is being evaluated.

Questions on relevance may cover the following areas:

¹⁷ This document draws from the OECD-DAC evaluation quality standards (<http://www.oecd.org/dac/evaluation/qualitystandardsfordevelopmentevaluation.htm>), the GRPP Sourcebook of the World Bank (<http://siteresources.worldbank.org/EXTGLOREGPARPROG/Resources/sourcebook.pdf>) and the CGIAR experience and guidelines for External Program and Management Reviews (http://ispc.cgiar.org/sites/default/files/ISPC_EPMR_Guidelines.pdf)

CGIAR Standards for Independent External Evaluation

- **Supply side relevance** relates to the alignment of the objectives of research and program components with the overall objectives of the CRP and its Intermediate Development Outcomes and the objectives and System-level Outcomes of the CGIAR.
- **Demand side relevance** relates to how well the program research and other activities correspond with the global, regional and national priorities and the needs of the targeted intermediate and final beneficiaries, differentiated by social groups.
- **Coherence** relates to the internal alignment and linkages between program components, and whether activities funded from different sources strategically complement each other.
- **Comparative advantage** relates to the CGIAR's role in delivering international public goods and complementing activities of global, regional and national partners, including the private sector, rather than competing or substituting for activities that other research institutions or development agencies can do more efficiently. It captures also the extent to which the research is responding to research opportunities to address agricultural development challenges. It also captures the complementary role that partners have in the program.

Program design assessment focuses on the clarity, logic, coherence and realism of the program and its different components and the program impact pathways and theories of change. It also examines the scientific relevance relative to research done elsewhere; balance of program components ranging from high-risk and long-term to applied, short-term research; consultation in designing or adjusting components of the program; realism in terms of cost and time; and prioritization.

Examples of evaluation questions:

- Is the program consistent with the main goals and System Level Outcomes of the CGIAR, and the agreed CGIAR reform agenda?
- Is there evidence of (continuing) demand for the program from intended beneficiaries?
- What is the comparative advantage of the program in terms of the CGIAR's mandate of delivering international public goods-relative to other international initiatives/research efforts, including the private sector; to partner country research institutions; and to development agencies?
- Is there a reasonable balance in the program portfolio between high and low risk activities; short-term and long-term research?
- Does the program plan to engage with appropriate partners, given their roles in implementation and achieving the objectives of the program?
- Have issues on social equity, gender and capacity been adequately addressed in program design and targeting?

Science Quality

All CGIAR evaluations with focus on research will need to formally assess the quality of science, not only because the CGIAR is a research organization, but also because its effectiveness depends on the science being of high quality. Therefore, also the System-wide evaluation and evaluations that focus on central CGIAR institutions, which have responsibilities related to research, will need to address science quality.

The ISPC has a vital role in assessing science quality in the CGIAR and promoting high science quality, in particular through appraising the science quality of CRP proposals.

Science quality refers to ways by which research is designed, conducted, documented and managed. Science quality is distinct from relevance although these aspects are often considered simultaneously. Evaluation of science quality looks both at the conditions that are in place for assuring high quality of science, and the conduct and outputs of research. A systematic and consistent evaluation of science quality across research programs and program components should have three dimensions:

- processes for assuring and enhancing science quality (staff recruitment, performance management and incentives; review processes used; codes of conduct; monitoring, evaluation and oversight for enhancing science quality);
- inputs (quality of staff and research leaders, facilities and equipment, data management, research design);
- outputs (volume and quality of publications, genetic materials, etc.).

As suggested in literature and practice in science quality evaluation, its evaluation should combine both quantitative and qualitative means of assessment. There are differences among the various fields of research in established research approaches, methodologies and documentation practices, which limit the comparability of different indicators used for science quality assessment. Therefore, evaluators should aim to determine acceptable levels of science quality across the program and in different areas of research, and to identify the components that meet these levels to those that do not. Competence on both research management and specific areas of research are needed in the evaluation team.

Examples of evaluation questions:

- Do the research design, problem-setting, and choice of approaches reflect high quality in scientific thinking, state-of-the-art knowledge and novelty in all areas of research?
- Is there evidence that the program builds on the latest scientific thinking and research results?
- Are the internal processes and conditions, including research staff and leadership quality, adequate for assuring science quality?

- Are the research outputs, such as publications, of high quality?

Efficiency

Efficiency is the extent to which the program has economically converted, or is expected to convert, its resources/inputs (such as funds, expertise, time, etc.) into results. In the research context, the assessment of efficiency refers to the activities and outputs that are under the control of the research program, and takes into account the exploratory nature and risks inherent to research. Cost-effectiveness is the extent to which the program has achieved, or is expected to achieve, its results at a lower cost compared with alternatives. Shortcomings in cost-effectiveness occur when the program is not the least-cost alternative approach to achieving the same or similar outputs, outcomes and impacts.

Given that the CGIAR deals with research where risk and serendipity are inherent characteristics, and the impart pathways from the new knowledge generated by research to development outcomes are, to a large extent, unpredictable and protracted, the evaluation of efficiency is realistically confined to the assessment of some aspects of organizational performance and operational efficiency.

A practical approach to evaluating efficiency is to assess the extent to which the organization or unit (CRP, for example) has good management and financial systems in place, administrative and transaction costs are efficiently managed, resources are allocated adequately and in a timely manner, partnerships, contracts, and, for example, competitive grants are efficiently managed. Accurate benchmarking with other organizations for assessing efficiency may be difficult, but evaluations should aim to detect obvious inefficiencies. To the extent possible, the evaluation will assess the program's cost-effectiveness, which takes the benefits arising from the program's activities as a given, and asks whether these could have been produced at a lower cost compared with alternatives.

An important efficiency-related question for the CGIAR, which should be addressed by all relevant evaluations (including all CRP evaluations), is the extent to which the CGIAR reform is resulting in increased efficiency of operations, including research management, partnerships and oversight.

Examples of evaluations questions:

- To what extent does the program have good financial management, budgeting, and reporting?
- Are facilities and services fully used? Is it worth considering outsourcing or amalgamating any services?
- How efficient are the criteria and processes for allocating the program's financial and human resources

- Is the level of collaboration and coordination with other CRPs appropriate and efficient for reaching maximum synergies and enhancing partner capacity?
- To what extent have the reformed CGIAR organizational structures and processes increased (or decreased) efficiency for successful program implementation?

Effectiveness

Effectiveness refers to the extent to which the program/project is making progress towards its objectives and is expected to achieve them, taking into account the exploratory nature and risks inherent to research. Assessing the effectiveness of research activities is a particular challenge, since research is an upstream activity that is often non-linearly and distantly linked to the Intermediate Development Outcomes of the program, and far removed from the CGIAR's System-level outcomes. Research may contribute only indirectly to changes in productivity, income, policy and other objectives, and achieving its objectives depends on capacity, investments and cross-sectoral development at national level, among other factors.

Evaluation will assess effectiveness by systematically reviewing the progress of the program's activities (outputs) in relation to plans, and the extent to which these outputs are contributing to the achievement of each activity's objectives (outcomes). It assesses the plausibility of a theory of change (or theories of change), the credibility and logic underpinning the assumptions and the extent to which the theory of change is adjusted following the accumulation of new evidence.

Evaluation explores the extent to which program design, management and monitoring enhance the likely achievement of program objectives; progress is demonstrated and documented, and learning takes place for adaptive management. Evaluation should also assess the extent to which specific conditions of different target groups are taken into account, and capacity needs and other enabling conditions addressed to increase the likely effectiveness of research.

Examples of evaluation questions:

- Has the CRP stayed on track in terms of progress and milestones toward outputs, and along the impact pathway toward outcomes?
- What were the major factors influencing the achievement or non-achievement of the planned and unplanned outputs and outcomes? Areas to consider may include, but are not limited to:
 - Quantity, quality and timeliness of inputs (including external funding).
 - Quality and timeliness of support to complementary programs and other contributions (such as high-level policy dialogue) expected from funders and other partners.

- Quality of science/research (see criterion above).
 - Quality of management and governance¹⁸, including effective use of monitoring.
 - Quality of partnerships and communication.
 - Mechanisms for and quality of feedback from intended users of the outputs and end beneficiaries.
 - Time frame: were plans realistic?
 - Management of risk: what risk monitoring and mitigation measures were used?
 - The role of different CGIAR institutions and the CGIAR reforms.
- The external environment. If external factors are identified as a major constraint, it should be investigated how this has been handled in the impact pathway: as an assumption, or as an issue to be tackled?
 - Is the monitoring system used effectively for adjusting the program on basis of lessons learned?
 - Have adequate constraint analyses and lessons from *ex post* studies informed program design for enhancing the likelihood of impact?
 - Is the CRP adequately addressing enabling factors for outscaling outcomes?
 - Has gender been adequately considered in CRP impact pathway analysis and implementation, understanding the differential roles of women and men along the impact pathway, generating equitable benefits for both women and men, and enhancing the overall likelihood of enhancing the livelihoods of women?
 - Are there any activities that should be modified, discontinued, or added to current portfolio in order to enhance the program's likely effectiveness?

Impact

In the IEA Glossary of evaluation terms impacts are defined as “Positive and negative, primary and secondary long-term effects resulting from a chain of events to which research has contributed, directly or indirectly, intended or unintended. These effects can be economic, socio-cultural, institutional, environmental, technological or of other types. “

Given that the pathway from research to impact is uncertain and protracted, and that long-term effects are very difficult, if not impossible, to attribute back to specific research programs or activities, these standards consider development outcomes and impacts interchangeably. They cover changes resulting from the use, adoption or influence of research outputs by the intended beneficiaries or changes in the status of natural resources.

¹⁸ [The GRPP Sourcebook](#) pp 71-86 provides guidance on assessing governance and management in global programs, including resource mobilisation and financial management

Evaluations covered by the IEA evaluation policy will not be able to directly assess impacts¹⁹. A practical approach for CGIAR evaluations is therefore based on meta-analysis through reviewing the impact and outcome assessments available from both monitoring and in-depth studies conducted by CRPs, Centre, SPIA or external impact assessors. Regarding evidence of impact, the ease or difficulty in attributing change to research interventions needs to be considered. In practice, adoption of technologies or policy influence can be useful proxies in the absence of credible studies of long-term or large-scale impacts and when it is not feasible to track long-term changes or attribute them to specific research interventions. Thus, relevant studies to be considered as evidence include also those on adoption and influence.

Examples of evaluation questions:

- What can be concluded from the findings of *ex post* studies? Aspects should include: the nature of the impact and distance from research; magnitude of impact in terms of area or beneficiaries; lessons regarding the conditions of constraints to impact; evidence of disaggregated effects on different groups in society, gender in particular; evidence of unintended positive or negative effects and their discriminatory nature regarding different groups.
- What was the basis for selecting research areas for *ex post* studies and is the coverage over program components or areas of research appropriate considering feasibility of impact assessment?
- Have the methodologies used been appropriate, taking the specific characteristics of the area of research into account? Where possible, evaluations should rely on verification procedures used in the CGIAR to validate outcome and impact studies.
- To what extent does the evidence of impact indicate the scale of impacts or further scaling taking place?

Sustainability

Sustainability is the continuation of benefits from an intervention after the work has been completed, and the probability of continued long-term benefits. Sustainability may be political, financial, institutional, economic, social and/or environmental. With research programs, evaluation of sustainability often depends on the evidence available for impact, hence these criteria overlap. Therefore, the evaluation questions on impact should encompass the sustainability of impact, although this adds difficulty to the evaluation.

This criterion also overlaps with evaluation of likely effectiveness, where the likely sustainability of results can be assessed. Aspects to evaluate include the extent to which

¹⁹ Main reason for this is that rigorous assessment of outcomes and impacts and their attribution to research is the provenance of specialized research studies and impact assessments that require specific skills, time and resources that are beyond what evaluations have.

enabling factors, such as capacity, gender-specific conditions or policies are taken into account. When sustainability is interpreted as related to operations and funding, it overlaps with the evaluation of efficiency and effectiveness of management.

Examples of questions specific to sustainability:

- To what extent are positive outcomes demonstrated at pilot or small-scale level likely to be sustained and outscalable?
- Is there an appropriate exit /handover strategy, for example in partnerships? Are partners prepared and incentivized to take on any necessary responsibilities?
- Have trade-offs between different longer-term outcomes been taken into account in program design and implementation, for instance regarding environmental sustainability?
- Do the program theories of change and the assumptions underlying these theories include aspects sustainability?