
Management position on the Comments of the Independent Office of Evaluation of IFAD on the IFAD11 Impact Assessment Report

Document: IFAD13/1/R.5/Add.2

Date: 17 February 2023

Distribution: Public

Original: English

FOR: REVIEW

Action: The Consultation on the Thirteenth Replenishment of IFAD's Resources is invited to review the Management position on the Comments of the Independent Office of Evaluation of IFAD on the IFAD11 Impact Assessment Report.

Technical questions:

Romina Cavatassi
Lead Economist
Research and Impact Assessment Division
e-mail: r.cavatassi@ifad.org

Aslihan Arslan
Senior Economist
Research and Impact Assessment Division
e-mail: a.arslan@ifad.org

Management position on the Comments of the Independent Office of Evaluation of IFAD on the IFAD11 Impact Assessment Report

Observations on the methodology of the IFAD11 IA	Management response
<p>1. IOE only reviewed the synthesis report of the IFAD11 IA. Within the time frame available, it was not feasible to review individual IAs prepared by RIA. IOE also notes that only one project that has been subjected to an IA was available for independent validation as of July 2022. For this reason, it was not possible to analyze the difference in evaluation ratings between projects with IA and projects without IA, as it was done for the IFAD10 IA.</p> <p>Compared to the IFAD10 IA, the IFAD11 IA displays several improvements, notably:</p> <ul style="list-style-type: none"> (i) an increased size of the sample of projects assessed (24 projects instead of 17 under IFAD10 IA, exceeding the initial target of 17 for IFAD11); (ii) the synthesis IFAD11 IA report presents an analysis of potential selection bias for the projects selected (Table 1, page 14); (iii) the presentation and discussion of the IFAD11 IA report are more measured than in the IFAD10 IA and some caveats are presented (page 17-18). Robustness checks are presented too (pages 19-21); (iv) the report highlights both area of strengths (the targets for improved income, improved productive capacities and increases market access were exceeded) as well as weaknesses (the target related to improved nutrition was not met). <p>The above makes the IFAD11 IA report clearer and more credible, compared with IFAD10. IOE recognizes the work done by Management for this IA and is aware of the methodological challenges and efforts that it entailed. At the same time, IOE wishes to share some caveats that will help appreciate the endeavour required for this IA, as well as the potential limitations and qualifications on the findings.</p>	<p>Thank you for these points underlining the achievements of the IFAD11 IA despite the challenging situations under which an unprecedented number of IAs have been finalized.</p> <p>Regarding the ratings from independent validations, we understand that this was an implication of the COVID-19 related delays and we invite IOE to re-engage in these independent validations as they find fit.</p> <p>We also invite IOE to examine sensitivity checks reported in Annex II that provide the basis for our conclusion that the meta-analysis results in the IFAD11 IA report are robust.</p>
<p>2. IOE suggests that the number of beneficiaries reported in the document should be taken with a degree of circumspection. This applies to the estimated total under IFAD 11 (112 million) and to the estimated numbers of beneficiaries experiencing improvements (pages 6-8). Data on project outreach are based, ultimately, on project M&E reporting. In IOE's experience, these are not fully reliable, due to:</p> <ul style="list-style-type: none"> (i) multiple-counting of the same beneficiaries (e.g., when a single beneficiary receives agricultural training support, credit, and free agriculture inputs, project statistics may report this as three beneficiaries); (ii) uneven project M&E reporting, 	<p>As IOE correctly notes, the number of beneficiaries that have achieved the threshold set in the RMF for income, production, market access, resilience and nutrition are based on project M&E systems that are indeed heterogeneous in terms of quality.</p> <p>Acknowledging the heterogeneity of details and quality of project M&E, it is also important to note that these are the numbers on which IFAD as well as other IFIs rely for general project outreach and reporting. To ensure conservative estimates, to the extent possible, it is useful to note that in the IA</p>

Observations on the methodology of the IFAD11 IA	Management response
<p>which often confuses between 'direct' beneficiaries and 'indirect' beneficiaries; and the latter tend to be estimated through generous assumptions.</p>	<p>report the "direct" beneficiaries are used in the projection rather than "indirect" beneficiaries that include the generous assumptions noted. The methodology has to rely on the reported numbers and the beneficiary numbers used are the best options available given the IFAD M&E system architecture.</p> <p>However we do think that it is important for there to be better quality of sample frames, and beneficiary recording in IFAD investments. This in turn is related to project design budgets and monitoring and measurement budgets that are held by OPR and RIA respectively, that are diminishing. Finally it is also important to regularly train M&E personnel in PMUs. This also has implications for IFAD's overall budgets. In this context, we also think that better monitoring and measurement systems at the start and ex ante (rather than at the end) is critical.</p>
<p>3. The criteria considered for the IA sample selection included items such as "buy-in from government and from IFAD" or "national or local conflict" (footnote 5). From a practical perspective, these are reasonable concerns but also pose risks of under-representing situations of weak performance. The IFAD11 IA ran ex-ante checks for possible sampling bias (table 1, page 14). This is an important step but some caution applies here as well. The tests of difference are conducted on indicators of project implementation progress, human/social capital/empowerment/environment and natural resources, and financial data. Financial data are drawn from IFAD databases: while the tests do not detect statistically significant differences, these are sizeable and may signal that the type of investments that they finance is different indeed.</p>	<p>Thank you for these comments. RIA would like to clarify few aspects.</p> <p>First and foremost, regarding conflicts it is worth noting that the level of conflicts considered as exclusion criteria are related to wars and real life threatening impediments in collecting data. For example, conducting impact assessment of projects that were supposed to close in Syria or Afghanistan was not considered feasible because it would mean the possibility of harm for IFAD personnel which would not be defensible.</p> <p>Secondly regarding country buy-in or IFAD engagement, it is important to clarify that this criteria is listed in the DEF as one selection criteria, which has been approved by the Board. It was included as a criteria primarily because as in all IFAD operations, approval from government is always required. It is important for sample selection to access secondary and monitoring data and this requires buy-in and engagement from governments. It is also important to note that in practice, this variable has not been a constraint in sample selection for Impact Assessments because IFAD has not encountered such a obstacle.</p> <p>In response to the last comment, that the sample selection criteria overall do not have implications for the final results reported. This is assured given that the IA teams have compared the overall distribution and attributes of the sample (sample values) with the overall population (the true values). This comparison shows that overall there is no ex-ante selection bias. Additionally as the IA report shows too (annex xx), to check for another level of robustness given the difference in the average values of some project characteristics (that are not statistically significant), a Heckman procedure was undertaken, that showed yet again that no selection</p>

Observations on the methodology of the IFAD11 IA	Management response
	bias was present and any forced "correction" would in fact have introduced bias (see Paragraph 4 in Annex 1). These steps combined address the worries about potential bias.
<p>4. Non-financial indicators are generated by supervision missions, based on project M&E systems. At IFAD, self- and independent evaluations have common findings about weak reliability of M&E systems and this needs to be considered. It would be useful to double-check the evolution of such indicators throughout the project life cycle and report on any changes or 'anomaly' observed. Also related to project selection, it is not clear whether the analysis controls for different project duration. Some projects may belong to a very different generation or be part of a multi-phase endeavour, challenging comparability.</p>	<p>Projects are usually notified of their selection into the IA sample during their last year of implementation (sometimes after project closure), therefore an endogenous improvement in M&E system quality that is reflected in ratings can be ruled out. The only possible impact of having been selected into the IA sample is seen in the difference of the t-tests of the rating on "Performance of M&E systems", which stems from the fact that the PCR quality improves when IA data are used. This does not reflect changes in performance during the project.</p> <p>The other comment about weak reliability of M&E is addressed in the comment 2.</p> <p>Regarding project duration: project duration is not one of the sample selection criteria for Impact Assessments. For IFAD11 IA, the sampled project's average duration is 8.4 years and the remaining 74 projects in the IFAD11 IA universe is 7.6 years, and the difference is not statistically significant at the 5% level. This indicates different project lengths do not have a significant impact on the reported results.</p>
<p>5. Also, related to the comparability, a question is whether any difference in country and local-level project context could be detected before the projects were launched. The impacts on the main indicators were estimated using pooled household data including country/project fixed effects. As there is only one project per country, country fixed-effect is the same as the project fixed-effect. It would be useful to clarify how this is interpreted.</p>	<p>The reported impacts on the main indicators were estimated through a meta-analysis. It is worth noting that within each IA, locally specific characteristics are taken into account in ensuring proper comparability between project beneficiaries and their counterfactuals.</p> <p>The fixed effects mentioned in the report refer to the pooled analysis that were conducted to corroborate the main findings, and take into account any unobserved variation that can stem from country/local political, institutional or social contexts. The observation is correct that these fixed effects would account for all project and country level such factors jointly, and cannot be separated into unobserved country characteristics vs. project characteristics given that there is one project per country in the pooled data. However this does not have implications for project results or for interpretation.</p>
<p>6. Meta analysis of diverse interventions. The IFAD11 IA summary, as in the past, is based on meta-analysis techniques. In the literature, a source of credibility for meta-analysis is the fact that the studies that are selected cover experiments that have more or less comparable objectives and are similar in nature and in measurement scale (e.g., a meta-analysis of assessments of conditional cash transfer programs to promote nutrition and health). In</p>	<p>All IFAD projects aim to improve the RMF indicators that are captured by the IAs and the meta-analysis. Therefore, all projects implemented by IFAD have a common philosophy and mandate that is implemented through a specific theory of change that can best achieve the overall development objectives. The specific indicators, within each project do differ depending on each Theory of</p>

Observations on the methodology of the IFAD11 IA	Management response
<p>contrast, the 24 projects in IFAD11 IA have diverse objectives and components. Thus, care needs to be applied when interpreting the aggregation of project-level impacts.</p>	<p>Change (such as crop production vs. livestock production), which is how each project contributes to the overall mandate and RMF tier II achievements. For the meta-analysis, as long as the odds ratios are represented in a way to indicate improvement in each indicator that has the same direction of change, the meta-analysis results are valid. It is precisely with this purpose in mind that the RMF tier II indicators used in the meta-analysis are fine-tuned based on each project's theory of change. In addition, specific results coming from project's logic and theory of change are reported in Impact Assessment briefs, reports and infographics.</p>
<p>7. In synthesis, the IFAD11 IA has built on the experience of IAs and, overall, has applied meta-analysis techniques in a judicious manner. There are still validity threats that are given by the nature of IFAD-funded projects and the diverse interventions that each project has supported, requiring prudent interpretation. Some caution is also needed before inferring that the changes observed can be solely attributed to the action of IFAD-funded projects.</p>	<p>The impacts presented in the IFAD11 IA report are averages, therefore cannot speak to the heterogeneity of project types. But the merit of doing meta-analysis also is that we know the distributions of these effects and overall their variance as well. This is why this work can help draw important design, delivery and overall results related recommendations for the institution, that individual projects analyses may not.</p> <p>Please also note that the IA report does not make claims of effects of individual IFAD interventions by various categories. To be able to report impacts by various categories would require a much larger IA sample that is stratified by each category. This is mentioned in the report (see paragraph 13 in Annex I).</p> <p>With regard to attributability of impacts, these are ensured through sample selection and in using rigorous quasi-experimental designs as described in the report. Potential contamination from other projects and interventions are always accounted for within each impact assessment.</p>

Observations from a developmental perspective	Management response
<p>8. It would be beneficial to contextualize the interpretation of findings. For instance, with reference to the finding that beneficiary income is, on average, 23 per cent higher than without IFAD investments (paragraph 10, Page 6), it would be helpful to specify whether this is nominal or in real terms and whether it refers to a point estimate or to the entire lifecycle of a project. Moreover, the practical importance of any given per cent change will depend on the starting conditions (it may be marginal in absolute terms if the starting point is low). As an example, the RIA IA of the Post-Tsunami Sustainable Livelihoods Programme for the Coastal Communities of Tamil Nadu (India) found that gross revenues from fish vending were 43 per cent higher for beneficiary</p>	<p>The reported percentage improvements are in real terms, as they compare beneficiaries with the counterfactual. They do not refer to changes over time that would require discounting for inflation & depreciation. All values are expressed in same terms, and indicate the difference between what beneficiaries have and what they would have had if they had not benefited from the project.</p> <p>The 23% average impact has been contextualized by using the GDP per capita of the 96 countries in the IFAD11 IA universe, and it has been established that this translates into a US\$225 increase in income per person per year (see paragraph 10 in the report).</p>

Observations from a developmental perspective	Management response
households than non-beneficiary ones. IOE conducted a Project Performance Evaluation of this project, noting that this was remarkable in relative terms but in fact the additional income (equivalent to US\$ 68 per year) made little difference in purchasing power, as explained by women beneficiaries during field interviews.	Considering the USD 1.95 per person per day international poverty line, the calculated average impact is not negligible for IFAD11 target beneficiaries (who on average have 30% of the national GDP per person given IFAD's targeting criteria and income distribution inequalities).
9. In addition, considering the COVID pandemic and the ensuing global economic crisis, it would be useful to qualify to what extent the +23 per cent reflects a net income increase for household with IFAD support or an income decrease for households without project or (probably) a combination of both.	Each IA that has collected household data after March 2020 included questions about the impact of COVID-19 on household incomes, output and input prices as well as other livelihood aspects. Moreover, the sample selection for each IA considered multiple indicators to ensure that the beneficiaries are comparable to control areas along multiple variables that can reasonably be expected to affect the exposure to COVID-19 (population density, distance to roads and cities, sector of income generation...etc) These variables have been used in statistical analyses to ascertain that the exposure to and effects of COVID-19 did not differ among beneficiaries vs. control. In a couple of cases where there were significant differences between the two groups, a composite COVID-19 exposure index was created and included in the impact analyses to control for differential impact. (Indeed this also speaks to some of the power of doing this kind of analyses.) Therefore this potential effect of the pandemic on the aggregate findings has been addressed to the extent possible within an ex-post quasi-experimental IA framework.
10. The definition of impact adopted in the IA reflects several but not all the domains of impact used at IFAD (with reference for example to the IFAD Evaluation Manual of 2015 or 2022). Notably it does not include impact on institutions and policies. This is hard to measure through surveys but important for rural development. If local or national institutions and policies can be made 'more pro-poor', they contribute to scaling up of project-level impacts and support sustainability of benefits.	RIA is totally in agreement with the importance and role of local and national institutions and policies to scale up impacts and strengthen development and sustainability of benefits. Rigorous measurements of these aspects, however, go beyond the scope of the Impact Assessment program of work. Corporate IFAD IA reports have to report on the RMF Tier II indicators. In IFAD11 IA report additional women's empowerment indicators have been presented to show progress made towards commitments on this mainstreaming theme. The institutional and policy perspectives cannot be captured in such a report using the established methodologies. The important role policies and institutions play in delivering impact can be captured by other types of studies that can be commissioned. The validity of the IA report does not hinge on this important aspect.
11. A broader perspective: blending IA with the review of other evaluation criteria. The international development evaluation practice is to adopt multiple evaluation criteria, so as to encompass several facets of development (OECD-DAC 2021). The impacts generated by a project would make limited change, if they are not sustainable. The IOE's Annual Report on Independent Evaluation (ARIE) of 2022 explores this aspect by computing the share of projects rated moderately satisfactory or better for rural poverty impact, first on its own and then combined with other selected criteria (sustainability,	Very relevant points. It is also important to note that in many cases, the sustainability of projects is reported by the projects themselves. IFAD IA reports do not report on sustainability given the commitments made under the IFAD RMF. Reporting on sustainability is a very important point which would require a different type of commitment, investments and analysis of which RIA is in full support but which goes beyond the scope of the Impact Assessment program of work and commitment.

Observations from a developmental perspective	Management response
<p>scaling up and efficiency). Out of the projects that were completed between 2011 and 2020, 81 per cent were rated by IOE as moderately satisfactory or above for impact. However, 60 per cent of projects were rated moderately satisfactory or above for both impact and sustainability (i.e., 21 per cent of projects were rated moderately satisfactory or higher for impact but moderately unsatisfactory or lower for sustainability). The key point is that, from a developmental perspective, impact is a central criterion but it is equally important to contextualize it and review it in conjunction with other criteria, to better qualify the findings.</p>	