

Document: EC 2020/109/W.P.4/Add.1  
Agenda: 5  
Date: 22 May 2020  
Distribution: Public  
Original: English

**E**



Investing in rural people

## **Comments of the Independent Office of Evaluation of IFAD on the Results of IFAD10 Sensitivity Analyses and Implications for IFAD11**

### **Note to Evaluation Committee members**

#### Focal points:

#### Technical questions:

**Fabrizio Felloni**  
Interim Officer-in-Charge  
Independent Office of Evaluation of IFAD  
Tel.: +39 06 5459 2361  
e-mail: f.felloni@ifad.org

#### Dispatch of documentation:

**Deirdre Mc Grenra**  
Chief  
Institutional Governance and Member  
Relations  
Tel.: +39 06 5459 2374  
e-mail: gb@ifad.org

Evaluation Committee — 109<sup>th</sup> Session  
Rome, 19 June 2020

---

**For: Review**

## **I. Background and general observations**

1. In September 2019, the Independent Office of Evaluation of IFAD (IOE) presented its comments to the Evaluation Committee and the Executive Board on the IFAD10 Impact Assessment Report. Further to the discussion, the Research and Impact Assessment Division of IFAD (RIA) prepared a sensitivity analysis on the same exercise.
2. The purpose of the present IOE comments is to offer an independent review of, and constructive insights on, RIA's analysis, "Results from IFAD10 Sensitivity Analyses and Implications for IFAD11", with a forward-looking perspective.
3. The Impact Assessment Initiative is a useful addition to IFAD's self-evaluation system. Impact assessments can be used by Management, as well as by IOE, as an additional source of evidence and information on portfolio performance. They can contribute to knowledge about performance and results and can support the preparation of new project designs.
4. IOE acknowledges that it is the first time that an international financial institution has tried to draw conclusions on the impact of its portfolio, based on a sample of projects. This is a challenging exercise both from the point of view of timely implementation and from an analytical perspective. The methodology for impact assessment was initially piloted under IFAD9 and further elaborated and implemented under IFAD10. It is now time to take stock of the experience.
5. IOE welcomes the effort made by RIA to conduct a sensitivity analysis, including the adoption of the Heckman correction for the sample selection model. The sensitivity analysis explains the type of data and information available when the IFAD10 Impact Assessment Initiative was launched. Most of the analysis is grounded in the project performance indicators available in 2016. These are indicators compiled annually in the project status report, based on project supervision mission findings.
6. A general observation from IOE is that, when evaluative findings are reported for accountability purposes, it is important to ensure that: (i) they are representative (i.e. they can be generalized to the entire population from which they have been extracted); (ii) limitations to their representativeness emerging before or after the impact survey are fully disclosed and explained.
7. Based on the sensitivity analysis and available data, IOE reiterates its appreciation of the exercise, while maintaining reservations as to the generalization of findings from a sample of assessed IFAD10 projects to the entire portfolio of projects completed under IFAD10. Key issues for attention are:
  - (i) The protocol for project selection focused on regional and thematic representativeness but did not include randomization procedures (e.g. lottery extraction) to control potential bias linked to purposive sampling;
  - (ii) While IOE recognizes the resource constraints, the size of the sample (19 projects under IFAD10) was small, threatening representativeness across the classes of project performance; and
  - (iii) Limitations and caveats to the generalization of the findings could have been further explained in the original report.

## **II. Specific observations**

8. As substantiated in the IOE comments of September 2019, an anomaly in the sample of the projects selected for the IFAD Impact Assessment was that their ratings across most evaluation criteria were significantly higher than those of other projects completed under IFAD10. This difference appeared both in the ratings of the project completion reports (prepared by IFAD Management) and in the ratings

of independent project-level evaluations (prepared by IOE). In a sample that is representative of the portfolio and balanced across the class of project implementation performance, such systematic differences would not be expected.

- (i) In the case of project completion reports, the RIA sensitivity analysis paper (page 12, table 1) confirms that average ratings were higher across all project-level criteria and the difference was significant ( $p$ -value  $< 0.05$ ) for 10 out of 14 project-level criteria;
  - (ii) As for IOE independent project ratings, the average ratings of projects selected for impact assessment were also significantly higher ( $p$ -value  $< 0.05$ ) for 11 out of 14 criteria: relevance, effectiveness, efficiency, project performance, impact on rural poverty, innovation, gender equality, adaptation to climate change, IFAD performance, government performance, overall project achievement (annex I, table 1 of the present document).
9. The sensitivity analysis concludes that, all the rest being equal, impact may have been overestimated by 2 per cent in the case of economic mobility, 15 per cent in the case of market access, 10 per cent for production, and 6 per cent for resilience, while nutrition results remain unchanged. Whether these are negligible differences is a question of interpretation and it would in any case be useful to consider the confidence intervals of these estimates. Many of the differences between estimates have a relatively large confidence interval, spanning up to 40 per cent for the Heckman-adjusted estimates.
  10. The analysis paper explores differences in the ratings of indicators from the project status reports available in 2016. It is important to note that these indicators: (i) tend to fluctuate across the project cycle and may not always be indicative of the final performance and results of a project; (ii) are not subjected to the same level of peer review and scrutiny as ratings of project completion reports or of independent evaluations.
  11. The analysis applies the Heckman sample selection correction method. A caveat is that, as shown in the literature,<sup>1</sup> Heckman's two-step approach may not provide correct estimates when sample sizes are small and there is high degree of multicollinearity between the estimated indicator value of the first equation and the right-hand-side variable in the second equation.
  12. The sensitivity analysis paper argues that the availability of impact assessment may have affected the ratings for impact on rural poverty. This should be taken with some caution: (i) projects selected for impact assessment had significantly higher average ratings for most criteria (not only for impact on rural poverty); (ii) in several instances (e.g. Brazil, Madagascar, Mexico, Sao Tome and Principe, the impact assessment was not available or finalized at the time of preparation of the project completion report and of the independent evaluation. In any case, even in the absence of an impact assessment, many projects had impact-level data from surveys, e.g. through the Results and Impact Management System or other surveys and studies commissioned by the project management unit.
  13. The distinction presented in the paper between "objective" and "subjective" indicators such as ratings should not be overstated. A large part of the sensitivity analysis paper is based on implementation performance indicators and these are ratings. Ratings are commonly used by self- and independent evaluations. Ratings are not arbitrarily assigned: they are set at the conclusion of an analysis of data and triangulation of evidence and are subjected to an internal peer review as well as a review by external stakeholders.

---

<sup>1</sup> Nawata, K. and Nagase N., 1996, Estimation of Sample Selection Models, *Econometric Review* 15, 387-400.

### III. Conclusions and way forward

14. IOE fully recognizes the practical and theoretical challenges to establishing a robust sampling strategy in the absence of satisfactory ex-ante performance indicators and with small sample sizes.
15. The sensitivity analysis provides clarity on the project performance indicators available when the IFAD10 Impact Assessment Initiative was planned. However, there are significant differences in the ratings between selected projects and those not selected, according to the self- and independent evaluations conducted after project completion. The reason for these ex post differences remains unexplained. This justifies a caveat before the Impact Assessment findings can be generalized to the entire IFAD portfolio.
16. To make a robust inference on the impact of the overall IFAD-funded portfolio, it is important to ensure that findings are representative and corrections are applied. Given the above observations, options that may be considered in similar future exercises include:
  - (i) Some form of stratified random sampling (e.g. via a lottery system applied to select the project sample). This may include some reserves for replacement (e.g. in case of an emergency situation), to be duly documented;
  - (ii) Broadening the sample size, tied to the level of precision required from the impact evaluation and expected variance in the distribution of key characteristics in the project population;
  - (iii) Weighting of project impacts in the meta-analysis (for example based on project completion reports or IOE ratings); and
  - (iv) Documenting limitations in the representativeness of the sample.
17. The IOE comments have not delved into other aspects of the Impact Assessment, such as the robustness of individual impact assessments and the strategy for the aggregation of findings. Some items that could be considered in the future include: (i) providing confidence intervals of the estimated impact; (ii) reviewing the assumptions made in the meta-analysis to aggregate findings; and (iii) explaining key features and possible caveats of the methodology applied for selecting samples in the individual project impact assessments. These would be valuable contributions not only for IFAD but also for other organizations that may wish to embark in a similar exercise in the future.

## Updated Comparison of ratings of Impact Assessment sample projects with available ratings from IOE project-level evaluations (in IFAD10)

Table 1

Comparison of IOE's average ratings of the Impact Assessment sample (19 projects) with other projects completed under IFAD10 (2016–2018) (88 projects)

| Indicator                                    | Average IOE ratings (IFAD10 IA sample) | Average IOE ratings (IFAD10 projects without IA) | Difference of average IOE ratings (IFAD10 IA vs IFAD10 no IA) | P-value t-test of average IOE ratings (IFAD10 IA vs no IA) | Median IOE ratings (IFAD10 IA sample) | Median IOE ratings (closed IFAD10 no IA) | Projects in IFAD10 IA sample with IOE rating greater than the median of non-IA projects (percentage) |
|--|--|--|---|--|---------------------------------------|--|--|
| Relevance                                    | 4.54546                                | 4  | 0.545455  | 0.01127*   | 5                                     | 4  | 55   |
| Effectiveness                                | 4.454545                               | 3.852941   | 0.601604  | 0.00664**  | 4                                     | 4  | 45   |
| Efficiency                                   | 4.18182                                | 3.55882  | 0.622994  | 0.03239*   | 4                                     | 3.5                                      | 82   |
| Sustainability                               | 4.09091                                | 3.70588  | 0.385027  | 0.192  | 4                                     | 4  | 36   |
| Project performance                          | 4.34546                                | 3.77206  | 0.573396  | 0.00201**  | 4.5                                   | 4  | 82   |
| Rural poverty Impact                         | 4.63636                                | 3.91177  | 0.724599  | 0.00095**  | 5                                     | 4  | 64   |
| Innovation                                   | 4.72727                                | 4.14706  | 0.580214  | 0.03337*   | 5                                     | 4  | 64   |
| Scaling-up                                   | 4.27273                                | 3.97059  | 0.302139  | 0.2548   | 4                                     | 4  | 36   |
| Gender equality and women's empowerment      | 4.54546                                | 3.91177  | 0.63369   | 0.02653*   | 4                                     | 4  | 45   |
| Environment and natural resources management | 4.27273                                | 4.125  | 0.147727  | 0.5843   | 4                                     | 4  | 45   |
| Adaptation to climate change                 | 4.45455                                | 3.92308  | 0.531468  | 0.0412*  | 5                                     | 4  | 55   |
| IFAD performance                             | 4.63636                                | 4  | 0.636364  | 0.00314**  | 5                                     | 4  | 64   |
| Government performance                       | 4.54546                                | 3.78125  | 0.764205  | 0.00725**  | 5                                     | 4  | 64   |
| Overall project achievement                  | 4.54546                                | 3.9375   | 0.607955  | 0.02214*   | 5                                     | 4  | 91   |

\* Significant at 5 per cent.

\*\* Significant at 1 per cent.

Source: IOE calculations based on the database of the Annual Report on Results and Impact of IFAD Operations (ARRI).