

Evaluation title	Impact Evaluation of WFP's Fresh Food Voucher Pilot Programme in Ethiopia, 10/2017-1/2019	
Evaluation category and type	Impact Evaluation	
Post Hoc Quality Assessment (PHQA) – overall rating	Satisfactory: 77%	
<p>This study assesses the introduction of the Fresh Food Voucher Pilot programme, an innovative nutrition intervention targeted at pregnant and lactating women in Ethiopia. The evaluation was designed as a cluster randomised controlled trial (RCT) with two implementation treatment arms. The study was well designed, with a rigorous methodology and with appropriate statistical power calculations. Nevertheless, the activity implementation was severely delayed, which meant that the evidence and findings from the evaluation are extremely limited. Given the implementation problems, it would have been advisable to extend the evaluation period in order to permit sufficient implementation coverage.</p>		
CRITERION 1: REPORT SUMMARY	Rating	Satisfactory: 84%
<p>The summary provides a very concise overview of the evaluation report. The main users of the evaluation are identified, all their names are listed, and all the key findings of the evaluation are summarised without errors or omission. The summary is upfront and transparent about the fact that its main objectives could not have been reached due to implementation delays. This section could have benefited from a more detailed description of the subject of the evaluation.</p>		
CRITERION 2: OVERVIEW OF EVALUATION SUBJECT	Rating	Satisfactory: 71%
<p>The overview section provides a concise and clear description of the targeted population, the activities, outputs and outcomes. The Fresh Food Voucher programme and the Social Behaviour Change Communication (SBCC) campaign are described as the subject of the evaluation. The report would have benefited from a more detailed discussion of the underlying mechanisms and Theory of Change to illustrate the key assumptions, mechanisms, and programme logic.</p>		
CRITERION 3: EVALUATION CONTEXT, PURPOSE, OBJECTIVES AND SCOPE	Rating	Highly Satisfactory: 90%
<p>The context is described in great detail and includes a lot of context knowledge on government nutrition priorities and other safety net programmes. The context of the evaluation is well explained, including background information on trends and indicators from numerous sources and various studies.</p>		
CRITERION 4: METHODOLOGY AND ANALYSIS	Rating	Satisfactory: 85%
<p>The evaluation was very well designed, with a planned RCT with two activity arms and a control group. It compared a voucher programme versus no voucher, as well as of two different intensity types. The RCT was planned with an appropriate sampling methodology and detailed design, including statistical power calculations with effect size analysis. However, the evaluation suffered as activity implementation was severely delayed which meant that the sample size proved to be too small for detecting reasonable effect sizes. The evaluation team should have considered changing the original design and control for baseline values rather than assessing changes over time. In addition, the evaluation did not incorporate a plan for a scenario where the implementation deviated from expected timelines.</p>		
CRITERION 5: FINDINGS AND RESULTS	Rating	Satisfactory: 75%
<p>Findings are transparently described and clearly present why the study hardly generated findings on outcomes due to delays in the implementation of the activities. The report also provides ex-post statistical power calculations, i.e. including the information about factual activity levels, in order to assess the statistical power of the impact evaluation after these operational changes. The overall presentation is well aligned with the planned methodology and carefully discusses strengths and weaknesses. The report would have benefited by including findings on the outputs in this section instead of presenting them in the "limitations" section, and thereby receiving less attention than they deserve.</p>		
CRITERION 6: CONCLUSIONS/LESSONS	Rating	Satisfactory: 79%

Conclusions are concise and transparent and follow the ordering of the evaluation questions. The conclusion section provides a comprehensive overview of the implementation fidelity and findings for the evaluation question. The conclusions provide a summary of the many evaluation challenges due to the implementation delays of the Social Behavioural Change Communication (SBCC) campaign and the Fresh Food Voucher (FVV) disbursement. They summarise the outcomes from the four evaluation questions.

CRITERION 7: RECOMMENDATIONS	Rating	Satisfactory: 75%
-------------------------------------	---------------	-------------------

Recommendations are relevant and concise. They are actionable and derived from earlier findings. However, due to implementation issues, recommendations are based on weak evidence. The recommendations mainly focus on implementation issues and propose several actionable suggestions related to the programme implementation.

CRITERION 8: ACCESSIBILITY AND CLARITY	Rating	Satisfactory: 74%
---	---------------	-------------------

The report uses tables and figures to illustrate the main results. Overall, the presentation is highly professional. All tables, figures and annexes are referenced and linked. The report uses visual aids such as tables and graphs to convey key information consistently throughout. However, the accessibility of the report could have been improved. For example, by moving the technical discussions to annexes as, at times, the language is too technical. The section on limitations is too long. As it contains very important information, this could have been a separate section describing activities and outputs and why it was not possible to conclusively answer the evaluation questions it set out to answer.

Integration of Gender Equality and Women’s Empowerment (GEWE) considerations in the evaluation report based on the UN System-Wide Action Plan (UN-SWAP) Evaluation Performance Indicator (EPI) scorecard

UN-SWAP EPI – individual evaluation score	Meets requirements: 8 points
--	------------------------------

The report meets the requirements and attains 8 out of 9 possible points. The project itself focuses on pregnant and lactating women, and all methods and evaluation questions take the situation of women into account. In addition, the analysis also examines boys and girls and any differences between them throughout.

Post Hoc Quality Assessment – Rating scale and definitions at overall report and criteria levels

Highly Satisfactory	<i>Definition at overall report level: Evaluation users can rely on the credible and useful evaluation findings provided and can use the evaluation with a high degree of confidence for decision-making. The report is considered an excellent example.</i>
	<i>Definition at criterion level: The criterion is addressed without any gaps or limitations.</i>
Satisfactory	<i>Definition at overall report level: Evaluation users can rely on the quality and credible evaluation findings provided and can use it with confidence for decision-making.</i>
	<i>Definition at criterion level: There are no significant gaps or limitations in addressing the criterion.</i>
Partly Satisfactory	<i>Definition at overall report level: Evaluation users can rely on the evaluation findings provided and may use it for decision-making noting that there are some gaps/shortcomings in the information provided.</i>
	<i>Definition at criterion level: There are some significant gaps or limitations in addressing the criterion.</i>
Unsatisfactory	<i>Definition at overall report level: Evaluation users can use some of the learning from the evaluation, noting that there are significant gaps/ shortcomings in the evaluation findings provided. The report may still contribute to decision making but should be used with caution.</i>
	<i>Definition at criterion level: There are critical gaps or limitations in addressing the criterion. Most of the required parameters are not met.</i>