

Evaluation title	Impact Evaluation of the Preschool Nutrition Pilot in Selected Counties of Xiangxi Prefecture, Hunan, PR China, 01/09/2018-19/11/2021	
Evaluation category and type	Impact Evaluation	
Post Hoc Quality Assessment (PHQA) – overall rating	Satisfactory: 79%	
<p>The Impact Evaluation of the Preschool Nutrition Pilot in Selected Counties of Xiangxi Prefecture, Hunan, PR China constitutes a satisfactory report that evaluation users can rely on the quality and credible evaluation findings provided and can use it with confidence for decision-making. The description of the subject is clear, with main activities, outputs and intended outcomes and targets are concisely described. The context provides mostly relevant information, illustrating how the pilot fits into and aims to complement the national policy and programmes on nutrition, poverty and agricultural development. The quasi-experimental control-group evaluation approach is well described and its limitations, vis-a-vis an experimental evaluation (RCT) design, are pointed out. Overall, the econometric analysis has been thoroughly conducted and the results have to be interpreted with care; however, limitations to the statistical precision for heterogeneity analysis and for outcomes where cluster-level correlations are likely could be made more explicit in the report. Findings are well presented and framed around the evaluation questions and sub-questions. The conclusions are largely balanced, comprehensively summarizes the main results. The report presents seven recommendations that are clear; however, several do not appear to flow logically from the findings and conclusions.</p>		
CRITERION 1: REPORT SUMMARY	Rating	Satisfactory
<p>Overall, a well written summary of the report and its findings and recommendations, presented without errors or omissions. The key elements are well summarized, outlining the methodology, including a detailed description of the household surveys, and limitations. While the overview explains the analytical basis of the subject, the subject is not described in detail and could have been further elaborated on. It would also have been useful to add the limitation of small sample size weighted clusters. Additionally, some of information provided in the report summary, specifically related to the context and subject, is not found in main report.</p>		
CRITERION 2: OVERVIEW OF EVALUATION SUBJECT	Rating	Satisfactory
<p>The main activities, outputs and intended outcomes and targets are concisely described. The description of the subject is clear, provides the key elements, and is based on relevant information sources. However, it is overly brief and would have benefited from more information about the program, some of which is provided in the summary but should have been repeated in the main report as well. The logic model is given, but only in the annex, with no discussion provided in the report. While an extensive list of references has been consulted and is displayed in an annex, the overview would also have benefitted from indicating the specific sources for particular information given.</p>		
CRITERION 3: EVALUATION CONTEXT, PURPOSE, OBJECTIVES AND SCOPE	Rating	Satisfactory
<p>The context provides mostly relevant information, illustrating how the pilot fits into and aims to complement the national policy and programmes on nutrition, poverty and agricultural development. The objectives of the evaluation are described. Information is provided on other projects that were implemented in the control group; however, the report only describes the situation before and at the beginning of the project and its evaluation but does not describe changes that happened during the evaluation. This is mentioned elsewhere in the report but should also have been indicated in the description of the context. Additionally, references of previous studies exploring similar programmes and evaluation question is not described or summarized. An intersectional analysis on the social groups affected is provided, but it is rather brief and quite limited in its scope.</p>		
CRITERION 4: METHODOLOGY	Rating	Satisfactory

The quasi-experimental control-group approach is well described and its limitations, vis-a-vis an experimental evaluation (RCT) design, are pointed out. Subsequent limitations to the design and econometric analysis are explained and mitigation strategies are clearly outlined. An evaluation matrix provides key element. A clear identification strategy of the counterfactual (control) is provided on how the impact and causal relationship is measured for each of the questions, as well as limitations. The sampling frame is explained in detail, with a clear rationale, and a detailed description of the data collection activities. The methodology permits analysis of effect heterogeneity at the individual level; however, a major limitation of the evaluation is the small number of cluster units due to the size of the project itself, which limits the extent to which treatment effect heterogeneity analysis is possible. While power calculations indicate reasonable minimum detection of effect size for nutrition outcomes (because of low intercluster correlation), for several other outcome variables (e.g. nutrition knowledge of preschool personnel or local farmers) larger cluster-level correlations are likely, and results could have been prone to spurious events due to small sample size. While power calculations and sample size issues clearly explained, these limitations could have been made more explicit.

CRITERION 5: FINDINGS AND ANALYSIS**Rating****Highly Satisfactory**

Findings are presented according to the econometric analyses by Evaluation Question and sub-questions, with a focus on learning about the causal mechanisms and intermediary channels, linked to the Theory of Change. Econometric methods and statistical inference are used correctly, with additional information given in tables in an annex. Findings and effect sizes are reported; however, "mean control group outcome" is missing for the subgroups analysed. Regression results are discussed throughout including any unexpected or surprising findings. The discussion particularly for EQ 1 is excellent, using graphical illustrations for the main results and descriptive tables, where the results are drawn from the detailed econometric annex. This is further triangulated with the results from the case studies. However, extensive treatment is given to the findings for EQ 1 at the expense of the other findings. On the other hand, the entire empirical strategy is targeted particularly towards verifying if the anticipated effects can be detected. This section could have been strengthened with a more extensive ex-ante discussion. Discussion on cost-effectiveness could have been further developed since the statistical analysis does not adjust for multiple-outcomes testing.

CRITERION 6: CONCLUSIONS/LESSONS**Rating****Satisfactory**

The conclusions are largely balanced, comprehensively summarizes the main results from the extensive findings across the three evaluation questions. There is no explicit assessment of replicability, but the conclusions indicate that an extension of the intervention should be possible, with the indication that the evaluation results are most relevant for poor areas in South Central China and not transferable to richer or poorer areas. Overall, the conclusions section is too short and would have benefited from further development, as several key details were left out, e.g. on the channels (i.e. lack of knowledge of caregivers). It would also have been useful to explicitly mention the sub-questions.

CRITERION 7: RECOMMENDATIONS**Rating****Satisfactory**

The report presents seven recommendations that are clear, taking into account contextual factors and WFP constraints. All recommendations are targeted to WFP and, in principle, WFP should be able to include these recommendations in further programming. Several recommendations do not appear to flow logically from the findings and conclusions, specifically recommendations 1 and 2. Most of the evaluation results were statistically insignificant (possibly due to small sample size) and some of the recommendations are based on rather weak evidence. Recommendations would have been strengthened by grouping them to enable sequencing/prioritisation by responsible actors.

CRITERION 8: ACCESSIBILITY AND CLARITY**Rating****Satisfactory**

Overall, the report is well written, presenting the main estimation results through the use of graphs and visual aids which lets the reader easily grasp the size of the estimated effects. On the other hand, the report is very long and would have benefited from streamlining. A few spelling mistakes and some informal wording could have been avoided to improve the quality of the summary.

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
<p>While gender and inequality considerations are not directly embedded in the evaluation questions, they are explicitly addressed throughout the findings. The evaluation places a strong emphasis on learning the effects for boys and girls and estimating heterogeneous effects by socioeconomic or poverty status of the household. GEWE is considered through the methodology and reflected in the data collection and analysis. Ethical standards are considered throughout the evaluation with reference to appropriate approval from Institutional Review Board (IRB). Evaluation methods and sampling frame included a diverse range of data sources and processes which addressed the diversity of stakeholders, primary data collection tool and sampling frame targeted to the poorer areas. The evaluation applies a wider concept of equality and incorporates economic inequality and vulnerability of children, e.g. children whose parents migrated for work-related reasons. However, greater attention could have been paid to the intersectional analysis the social groups affected, which is quite brief and limited in scope, and assessment of monitoring data, which is largely missing.</p>	

Post Hoc Quality Assessment – Rating scale and definitions at overall report and criteria levels	
Highly Satisfactory	<i>Definition at overall report level:</i> 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>Definition at criterion level:</i> The criterion is addressed without any gaps or limitations.
Satisfactory	<i>Definition at overall report level:</i> Evaluation users can rely on the quality and credible evaluation findings provided and can use it with confidence for decision-making.
	<i>Definition at criterion level:</i> There are no significant gaps or limitations in addressing the criterion.
Partly Satisfactory	<i>Definition at overall report level:</i> 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>Definition at criterion level:</i> There are some significant gaps or limitations in addressing the criterion.
Unsatisfactory	<i>Definition at overall report level:</i> 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>Definition at criterion level:</i> There are critical gaps or limitations in addressing the criterion. Most of the required parameters are not met.