

Field Trial of Local Nutrition Plans and Programs Monitoring and Evaluation Protocol in the Philippines

Leila S. Africa^{1*}, Nancy A. Tandang², Ma. Theresa M. Talavera¹, Nelson Jose Vincent B. Querijero³, Wilfredo B. Carada³, Kristine V. Montecillo^{1,4}, Angelina R. Bustos¹, Aileen R. De Juras¹, Mayo Grace C. Amit³, Hygeia Ceres Catalina B. Gawe⁵, Jasmine Anne F. Tandingan⁵

¹College of Human Ecology, University of the Philippines Los Baños, College, Laguna 4031, Philippines

²College of Arts and Sciences, University of the Philippines Los Baños, College, Laguna 4031, Philippines

³College of Public Affairs and Development, University of the Philippines Los Baños, College, Laguna 4031, Philippines

⁴Science Education Institute, Department of Science and Technology, Taguig City, Metro Manila 1631, Philippines

⁵Nutrition Surveillance Division, National Nutrition Council, Taguig City, Metro Manila 1231, Philippines

ABSTRACT

The field trial was conducted to establish the reliability in producing similar results between evaluators of the proposed new tools for Monitoring and Evaluation (M&E) of the nutrition plans and programs in the Local Government Units (LGUs). To do this, orientation activities were conducted to familiarize the 46 M&E team (MET) members evaluating the provincial, municipal, city, and barangay levels in two regions with the proposed tools during the field trial. After the event, the perceptions of the MET members of the tools were gathered by asking them to rate the tools through a self-administered questionnaire, and by noting their written and verbal commentaries about the proposed system. During the field trial, each MET member, as well as the member of the Project Team (PT), individually evaluated the LGUs using the tools. Secondary data on the LGUs performance scores using the old system were also gathered. The MET members' perception was examined based on the median rank of their ratings and content analysis of their insights about the tools, whereas the reliability of the tools was assessed based on the interrater reliability of the MET members' scores for the LGUs analyzed through paired samples t-Test, Pearson correlation coefficient, intraclass correlation coefficient, and technical error of measurement. The weighted scores of the MET and PT members were also compared. Moreover, the difference in the generated scores between the old and the new system was determined. The findings revealed that the MET members generally have a positive perception of the new system but raised some issues and concerns. Although the reliability of the tools was generally observed, actions are warranted for improvement. The tools generated statistically different scores when used by MET and PT members, and when compared to the existing system. Steps should be taken to improve the reliability of the proposed tools.

Keywords: interrater reliability, LGNMES, MELLPI Pro, monitoring and evaluation system, nutrition plans and programs

INTRODUCTION

Mandated to monitor and evaluate the implementation of the Philippine local food and nutrition plans and programs, the National Nutrition Council (NNC) has been implementing the Monitoring and Evaluation of Local Level Plan Implementation (MELLPI) system since 1978. This was employed to assess the efficiency and effectiveness of the LGUs in planning and

implementing local nutrition programs. However, in 2015, the NNC initiated the enhancement of the system to cover the assessment of the local nutrition policy and legislation initiatives, service delivery, and capacity-building, and include nutrition outcomes among pregnant mothers, in addition to infants and young children to make mobilization for nutrition more facilitative for LGUs. Consequently, an updated M&E protocol with the appropriate tools was developed

*Corresponding Author: tel: +63495362445, email: lsafrica@up.edu.ph

(Received 08-12-2020; Accepted 17-06-2021; Published 29-07-2021)

adapting a more results-based, integrated, and comprehensible approach (Gawe 2015) and is anchored on the various hierarchies of nutrition action plans. The updated M&E protocol was based on the Local Government Nutrition Monitoring and Evaluation System (LGNMES) of the Nutrition Results Framework (NRF) that is developed by the University of the Philippines Los Baños (UPLB). One of its components is the Local Nutrition Organizational Capacity Assessment Component (LNOCAC) which assesses the nutrition plans and programs of the LGUs.

The updated M&E system was pre-tested for feasibility, relevance, and comprehensiveness but a field trial was needed to assess the reliability and replicability of the tools. A field trial can provide relevant information before making any public health decisions and optimizing national health programs (Piedra-Fernández & Ganoza-Guerrero 2016). It establishes whether a change in the system can lead to a more desirable outcome ensuring that the new protocol will not fail or even worsen the existing one before its wide-range implementation (Smith *et al.* 2015; Wandner 2017). Furthermore, since the M&E information generated would be the basis for nutrition-specific planning, intervention, and policymaking (Vidyarini *et al.* 2021; Jefferds & Flores-Ayala 2016), it is vital to ensure that data produced are reliable and reproducible over time (George *et al.* 2013). To do this, the interrater reliability of the personnel in using the M&E tools must be established to assure consistency in the evaluation of a particular object or event (Drummond & Murphy-Reyes 2017).

Hence, this study aimed to conduct a field trial of the LNOCAC of LGNMES in the M&E of nutrition plans and programs at the different levels of LGUs. Specifically, it aimed to report the perceptions of the evaluators about the proposed protocol. It also assessed the reliability of the new system in evaluating the local nutrition plans and programs. Lastly, it determined the implication of the adoption of the proposed system in the performance evaluation of the LGUs.

METHODS

Design, location, and time

Two regions were selected jointly by NNC and UPLB as the study areas for the field trial. The

criteria used for the selection of provinces, cities, municipalities, and barangays were as follows: has records of 2016 MELLPI scores; with active and organized local Nutrition Plans and Programs (NPP) evaluators or monitoring and evaluation team (MET); Local Chief Executive's (LCE) approval of participation; and willingness of the MET members to be part of the study. Using these criteria, the NNC Regional Offices were consulted for the identification of the study areas as well as participants. The field trial was conducted from April to May 2018 in four provinces, two cities, and eight municipalities in the two regions.

Sampling

Upon coordinating with NNC Regional Offices, 48 MET members qualified and were thereby invited to participate in the study. The qualified MET members were those active and organized evaluators of local NPP in the selected areas identified by the NNC Regional Offices. Among the invited MET members, 46 agreed to participate. The selected MET members evaluated the LGUs corresponding to their levels of M&E protocol using the LNOCAC tools during the field trial: the Regional M&E Team (RMET) was assigned to evaluate a province or a city; the Provincial M&E Team (PMET) was assigned to a municipality; and the City/Municipality M&E Team (C/MMET) was assigned to a barangay. A total of six RMET members, 11 PMET members, six CMET members, and 23 MMET members participated in the study. All the MET members were asked to sign an informed consent stating their willingness to participate in the study.

Data collection

Before the field trial, the Project Team (PT) conducted activities to orient the MET members about the new LGNMES tools and protocol. These activities included the evaluation of LGUs corresponding to their levels of M&E protocol using the LNOCAC of the LGNMES, the focus of this paper.

After the orientation activities, the MET members were asked to answer a self-administered questionnaire about their perception of the LNOCAC protocol. Participants accomplished the questionnaire by expressing their agreement or disagreement with each statement based on a 5-point Likert scale. Each item was rated on a 1 to 5 response scale where; 1=strongly

disagree; 2=disagree; 3=agree; 4=strongly agree; and 5=neither agree nor disagree. The said questionnaire captured the whole aspects of understanding and using the evaluation tools. It inquired about the understandability of the words used in the tools, the scoring system, easiness in interpreting the results, easiness in explaining to Local Nutrition Committees (LNCs)/LCEs, its relevance, comprehensiveness, clarity, applicability to the full range of intended uses, concreteness, parsimony, ease of use, and if fairness were integrated. The MET members' written remarks and verbal commentaries during the orientation activities were also noted.

Following the LGNMES protocol, the METs were asked to do a certain sequence of activities during the field trial: travel to the site or venue of evaluation; conduct of courtesy call to the LNCs and orientation; desk review and scoring the LGUs' performances using the LNOCAC tools/forms; the processing of the team; and feedbacking of results to the LNCs. The LGUs were requested to make the relevant documents available during the scheduled MET visit to facilitate the evaluation process.

During the scoring process, each MET member individually evaluated the LGUs corresponding to their level of M&E protocols. This was done to determine the Interrater Reliability (IRR) between the MET members in using LNOCAC forms of the LGNMES for evaluating LGUs. A PT member assigned to the area also evaluated the LGUs using the tool to serve as a source of comparison. Both teams scored the performance of the LGUs based on the following dimensions in the LNOCAC forms/tools: Vision and Mission, Nutrition Laws and Policies, Governance and Organizational Structure, Local Nutrition Communication Management Functions, and Nutrition Intervention under its Organizational Component, and Prevalence of Underweight 0-<5 Children, Prevalence of Stunted 0-<5 Children, Prevalence of Wasted 0-<5 Children, Prevalence of Overweight and Obesity among Children, Prevalence of Wasted School-Age Children, and Prevalence of Nutritionally At-risk Pregnant Women under its Nutrition Situation Component.

Lastly, the data on the 2016 MELLPI scores of the LGUs were also gathered during the visit. These were collected to determine the implication of the adaption of the LGNMES in

the performance scores of LGUs. This is vital to evaluate the two systems based on the generated performance scores (Figure 1).

Data analysis

Perception of the MET members. The data on the participants' perceptions were gathered to determine their view of the proposed system and the possible implication of its implementation in different levels of LGUs. For the analysis of the perception of the MET members on the LNOCAC tools and protocol, the median rank of their ratings per item was examined by region. The median rank of ratings in the two regions was also statistically compared using the Mann-Whitney U-test at a 5% level of significance. The common themes in the written remarks in the questionnaire and verbal commentaries during the orientation activities of the participants were also determined through content analysis.

Reliability of the LNOCAC tool. Reliability is an important indicator to assess the usefulness of a tool. In this study, the reliability of the proposed tool in monitoring and evaluating the performance of the LGUs was estimated to describe the capacity of the said tool to produce almost similar results (precision) across several evaluators. To assess the reliability of the LNOCAC forms in evaluating the LGUs' performance, the IRR of the MET members based on their scores in using the tools were analyzed through t-test for paired samples, Pearson correlation coefficient, Intraclass Correlation Coefficient (ICC), and Technical Error of Measurement (TEM). The weighted mean performance scores of the MET and PT members for the different levels of LGUs were also compared to describe the precision or capacity of the tools to produce precise results when used by different groups of evaluators.

The paired t-test was used to describe the agreement between two MET members' performance scores for an LGU. A $p > 0.05$ indicates that the scores given by the two MET members are in agreement. The Pearson correlation coefficient, on the other hand, measured the degree of association between two MET members' performance scores for an LGU. A positive Pearson correlation coefficient indicates a direct association in the scores of the two MET members, while a negative coefficient indicates an opposite direction implying poor association. The closer its value is to one, the

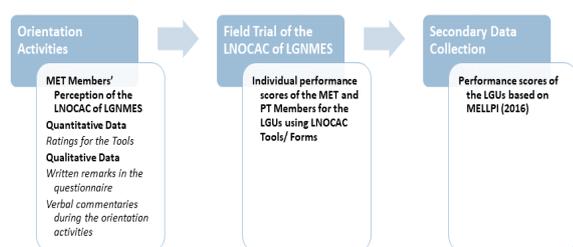


Figure 1. Flowchart summarizing the information gathered in each phase of data collection

stronger the association of the scores. Moreover, the ICC was computed to evaluate the degree of reliability of the two MET members' performance scores for an LGU: a negative coefficient indicates no reliability; coefficients less than 0.5 are indicative of poor reliability; 0.5 to <0.75 is moderate reliability; 0.75 to 0.90 is good reliability; and coefficient more than 0.90 indicates excellent reliability. It reflected both degrees of correlation and agreement between measurements, and thus, was used as an index. Lastly, the TEM was computed to indicate the IRR of the MET members in the absence of paired t-test, Pearson correlation coefficient, and ICC results due to constant scores between two evaluators. The TEM was used to indicate the variability between the two MET members' performance scores for an LGU. The lower the computed value, the more the two evaluators are in agreement. The percentage distributions of the acceptable IRR of the MET members by dimensions of the LNOCAC were then computed to describe the reliability of the tools.

In comparing the weighted mean performance scores of the MET and PT members for the different levels of LGUs, a significant differentiation of their scores indicates poor reliability of the tools. This comparison of the scores was also based on the paired t-test, Pearson correlation coefficient, ICC, and TEM.

Comparison of M&E systems. The LGUs' weighted mean performance scores based on the LNOCAC tools as evaluated by the MET and PT members were compared with the corresponding 2016 MELLPI scores. The weights were assigned to generate comparable performance scores. The comparison was done to determine the possible implication in the performance scores of LGUs upon the adaption of the LNOCAC as the new M&E system of the implementation of NPP at the local levels in the country. The paired t-test was

used to determine the differences between the weighted mean performance scores for the LGUs based on the two M&E systems. The data on this study were analyzed using the Statistical Package for Social Science (SPSS) version 23.

RESULTS AND DISCUSSION

MET members' perceptions of the new system

Table 1 shows the perception of the MET members on the LNOCAC as an M&E system for the NPP of the LGUs. The MET members in Region B generally had higher median ranks of ratings on positive statements about the new system than in Region A. This indicates that MET members in Region B agree to these statements more than those in Region A, albeit statistically insignificant.

On the other hand, the results of the median rank of ratings on the negative statements that compare the LNOCAC to MELLPI indicate that the proposed protocol takes less effort to use and that it is preferred in the M&E of the nutrition situation the LGUs in both regions. Moreover, the MET members in both regions had median ranks of ratings of three in the "bored on the system of evaluation" statements indicating the general agreement to the statement. Significant differences in the median ranks of ratings given by MET members in the two regions were only observed in the statements "The evaluation is costly." ($p=0.013$) and "It is too long to finish the evaluation." ($p=0.001$). The result in the statements about the costliness indicates that MET members in Region A agree that the LNOCAC is costly while MET members in Region B believe otherwise. On the other hand, the evaluators in both regions disagreed on the statement regarding the duration of the process, although MET members in Region B had a significantly stronger disagreement with the said statement compared to Region A.

Furthermore, based on the content analysis of the MET members' written remarks and verbal commentaries during the orientation activities, the LNOCAC is more comprehensive compared to MELLPI because of its broad range of organizational dimensions while also considering the changes in nutritional status as an indicator of nutrition programs' effectiveness like MELLPI. However, they also highlighted that the proposed tools were designed to be used only to monitor

Field trial of the LGNMES in the Philippines

Table 1. The median rank of the ratings given by MET members for the LGNMES in evaluating LGUs

Perceptions	Region A	Region B	p-value
Positive statements			
The evaluation is relevant to improve the nutrition situation in the LGU	3	4	0.352
The evaluation is comprehensive	3	3	0.154
The evaluation will be useful for the LGU	3	4	0.206
The evaluation will be useful for the LCE	3	4	0.298
The evaluation will be useful for the LNC	3	4	0.212
The evaluation will be useful for the nutrition workers in the LGU	3	4	0.358
The words used in the form are easy to understand.	3	3.25	0.181
The criteria are easy to interpret	3	3	0.904
The form is easy to fill out	3	3.5	0.824
The instructions on how to use the form are clear	3	3	0.993
The criteria used is fair from LGU to LGU	3	3	0.201
The results are easy to interpret	4	4	0.703
The results are easy to explain to the LNC	3	3.5	0.675
I enjoy using and filling out the forms	3	3	0.939
Training is not needed to use in this evaluation system, orientation is enough	3	3	0.802
Negative statements			
The evaluation is costly	3	2	0.013*
It is too long to finish the evaluation	2.5	1	0.001*
It takes more effort to use this evaluation system than MELLPI	2	2	0.061
I prefer MELLPI over this new tool	2	2.25	0.941
I am bored with this system of evaluation	3	3	0.029

*Significant at 5% level; MET: Monitoring and Evaluation Team; LGNMES: Local Government Nutrition Monitoring and Evaluation System; LGU: Local Government Units; LCE: Local Chief Executive's; LNC: Local Nutrition Committees; MELLPI: Monitoring and Evaluation of Local Level Plan Implementation

and evaluate local NPP implementation while the results of the MELLPI system are used as a basis for giving awards or recognition.

The MET members suggested that perhaps the MELLPI is still needed to be used in tandem with the proposed tool. The two tools may be harmonized by either merging their important elements or by simply making the proposed tool a continuation of MELLPI. They added that it may be necessary to retain the data validation processes and the presence of the Barangay Nutrition Council (BNC) members protocol during the evaluation. Moreover, they mentioned that there is a need to synchronize the LNOCAC forms with other health-related forms being used by LGUs. They also claimed that the Department of the Interior and Local Government (DILG), the executive department of the national government

for strengthening LGUs, should be part of the MET so that it will have a strong impact.

The participants identified some factors that needed to be considered before the nationwide implementation of the LGNMES. They say that the proposed system may require added manpower and better logistics implementation from the LGUs. The tools must also be further improved as it includes parameters that did not apply to all LGUs such as protracted disaster areas, but does not consider ordinances related to health, solid waste, and sanitation which are relevant to the overall nutrition situation of the communities. Additionally, the MET members deemed that the answer for minimum change in nutritional status could be manipulated and that the new tools are costly due to their numerous pages.

Moreover, they mentioned that the tools would be much better if it was shorter and more specific to the area being evaluated. The criteria in each dimension should also be improved by identifying a basis for comparison or which data to use. Lastly, contrary to the overall result of the self-administered assessment of the LNOCAC system, some MET members still think that intensive orientation training should be provided.

Reliability of the LNOCAC tools

Interrater reliability of the MET members.

The results in the t-test for paired evaluators revealed that all MET members had a hundred percentages of acceptable IRR, except for the Nutrition Intervention dimension among MMET which had 80%, indicating their high degree of agreement with each other in their scores for the LGUs in all the dimensions of the LNOCAC tools. The Nutrition Situation Component of the tools had generally better reliability than its Organizational Component as lower percentages of acceptable IRR among MET members were observed in the latter. The reliability of the tools was also more evident in Region A based on the higher percentages of acceptable IRR observed among MET members than in Region B.

Furthermore, the results showed that the RMET members had better IRR with one another in using the LNOCAC tools than in other groups. This indicates that the tools used for the M&E of NPP of the provinces and cities are more reliable than the LNOCAC tools used in other levels of LGUs. The Nutrition Situation Component of the tools for the barangays in cities also showed notable reliability based on the high percentages of acceptable IRR of the CMET members (Table 2).

Moreover, the Nutrition Intervention dimension under the Organizational Component had 100% percentages of acceptable IRR consistently among the MET members based on ICC while high percentages of acceptable IRR among the MET members were commonly observed in Prevalence of Overweight and Obesity among Children and Prevalence of Wasted School-Age Children dimensions under the Nutrition Situation Component. The lowest percentages of acceptable IRR among the MET members based on ICC were generally recorded in the Vision and Mission, Nutrition Laws and Policies, and Local Nutrition Communication Management Functions dimensions under the

Organizational Component. On the other hand, the MET members also had the lowest percentages of acceptable IRR in the dimension under the Nutrition Situation Component, particularly in using the tools for the M&E of provinces, cities, and municipalities NPP. These results indicate that the LNOCAC tools for evaluating local NPP are least reliable in these dimensions.

The reliability of the LNOCAC tools when used by different groups. The results of the comparison of the MET and PT members' weighted mean performance scores for the LGUs using the LNOCAC tools are summarized in Table 3. Based on the paired t-test analysis, an agreement was observed between the MET and PT members based on their weighted mean performance scores for the LGUs, except for the barangay levels in Region A ($p=0.003$).

Excellent IRR was observed between the two groups based on their weighted mean performance scores for the provinces in Region B as evidenced by the computed value of ICC. This indicates the reliability of the LNOCAC tool for evaluating NPP at the provincial levels when used by different groups of evaluators. The two groups also had the smallest differences in the scores when the tool was used in these areas based on the TEM. However, contradicting results were found based on the tool when used by the two groups in Region A; the mean performance scores of the two groups for the provinces had a negative correlation and ICC. The highest variability in the scores was also observed when the tool was in these areas based on the TEM.

Moreover, the two groups had moderate reliability in using the LNOCAC tools for evaluating the barangays in both regions and municipalities in Region B, while poor reliability for evaluating the municipalities in Region A and cities in Region B based on ICC. To sum, the results showed that the LNOCAC tools were reliable in varying degrees when used by different groups for evaluating the NPP of different levels of LGUs, except for provinces in Region A when the tools were found to have no reliability based on the ICC.

Comparison of LGNMES and MELLPI

The adjusted mean performance scores of both the MET and PT members using the LNOCAC tool were compared to the corresponding 2016 MELLPI scores of the LGUs

Field trial of the LGNMES in the Philippines

Table 2. Percentage distribution of acceptable interrater reliability of MET members

Dimension	Region A			Region B			Region A			Region B		
	Corr coeff	ICC	TEM	Corr coeff	ICC	TEM	Corr coeff	ICC	TEM	Corr coeff	ICC	TEM
MMET						PMET						
<i>Organizational component</i>												
VM	80	80	70	83.3	83.3	66.7	83.3	50	66.7	100	100	75
NLP	70	60	100	83.3	66.7	50	100	66.7	66.7	50	50	50
GOS	80	70	80	66.7	58.3	50	100	83.3	50	25	75	25
LNCMF	90	70	70	58.3	41.7	75	83.3	83.3	33.3	50	50	66.7
NI	100	100	60	75	66.7	91.7	100	100	16.7	100	100	100
<i>Nutrition situation component</i>												
PUW	90	90	70	83.3	83.3	66.7	66.7	66.7	66.7	100	100	75
PS	100	100	50	66.7	83.3	50	100	100	66.7	75	75	75
PW	90	90	70	91.7	91.7	50	66.7	66.7	33.3	100	100	75
POO	80	80	60	75	66.7	50	100	100	16.7	100	100	75
PWS	80	80	60	75	75	50	100	100	33.3	100	100	75
PNAP	80	80	70	83.3	83.8	75	100	100	50	75	75	75
CMET						RMET						
<i>Organizational component</i>												
VM	100	100	33.3	33.3	33.3	0	100	100	100	100	100	33.3
NLP	100	100	100	33.3	33.3	33.3	100	100	100	100	100	100
GOS	100	100	66.7	66.7	66.7	0	66.7	66.7	0	100	100	100
LNCMF	100	66.7	33.3	33.3	0	0	33.3	33.4	33.3	100	100	100
NI	100	100	100	100	66.7	0	100	100	100	100	100	100
<i>Nutrition situation component</i>												
PUW	100	100	100	100	100	100	100	66.7	100	100	100	100
PS	100	100	100	100	100	100	100	100	100	100	75	100
PW	100	100	100	100	100	0	100	66.7	100	100	100	100
POO	100	100	100	100	100	33.3	100	100	33.3	100	100	100
PWS	100	100	100	100	100	33.3	100	100	33.3	100	100	100
PNAP	100	100	100	100	100	33.3	100	100	100	100	75	100

GOS: Governance and Organizational Structure; ICC: Intraclass Correlation Coefficient; LNCMF: Local Nutrition Communication Management Functions; NI: Nutrition Intervention; NLP: Nutrition Laws and Policies; PNAP : Prevalence of Nutritionally At-risk Pregnant Women; POO: Prevalence of Overweight and Obesity among Children; PS: Prevalence of Stunted 0-<5 Children; PUW: Prevalence of Underweight 0-<5 Children; PW: Prevalence of Wasted 0-<5 Children; PWS: Prevalence of Wasted School-Age Children; TEM: Technical Error of Measurement; VM: Vision and Mission; MMET: Municipality Monitoring and Evaluation Team CMET: City Monitoring and Evaluation Team; PMET: Provincial Monitoring and Evaluation Team; RMET: Regional Monitoring and Evaluation Team

to implications of the adaption of the LGNMES as the new M&E system for evaluating the local NPP (Table 4). Results showed that the MET and PT members' adjusted scores were consistently lower than the 2016 MELLPI scores across all the LGUs evaluated. However, these differences were only found significant for the performance scores of barangays in both regions, and the city in Region B.

Furthermore, significant differences were also observed between the 2016 MELLPI weighted score and PMET members' mean

performance score for the municipalities, and between the 2016 MELLPI weighted score and PT members' mean performance score for the provinces in Region A. No significant differences were observed in the performance scores for municipalities and provinces in Region B. The overall results indicate that LGUs' performance scores would significantly become lower when the LGNMES tool is adapted as the new M&E system, particularly at the barangay levels.

The MET members had a generally positive perspective on the LNOCAC of LGNMES,

Table 3. Comparison of the MET vs PT members' consensus scores using the LGNMES tool

	Mean Scores (Standard deviation)		p-value (t-test) ^a	Corr coeff	ICC	TEM
	MET Members	PT Members				
Region A						
Barangay	51.78 (17.54)	42.04 (16.51)	0.003*	0.805	0.699	98.35
Municipality	60.61 (22.79)	65.80 (20.85)	0.664	0.357	0.396	259.3
City ^b	44.37	48.82	-	-	-	9.9
Province	55.35 (12.32)	38.19 (16.85)	0.412	-0.966	-0.890	426.2
Region B						
Barangay	61.35 (17.08)	51.03 (21.42)	0.059	0.796	0.697	112.6
Municipality	72.02 (15.92)	60.95 (17.15)	0.070	0.749	0.632	119.0
City ^b	58.40 (14.53)	44.78 (17.12)	0.151	0.608	0.473	168.3
Province	55.54 (3.83)	54.72 (3.95)	0.065	1.000	0.978	1.4

*Significant at 5% level; a) using t-test for paired samples; b) with one observation only; MET: Monitoring and Evaluation Team; PT: Project Team; ICC: Intraclass Correlation Coefficient; TEM: Technical Error of Measurement

the proposed M&E system for evaluating local NPP. However, they raised some concerns such as implied logistics challenges, additional manpower requirements, cost implications, and technical issues about the tools that need to be resolved and considered prior to its nationwide implementation. The reliability of the tools was generally observed but is more evident in the Nutrition Situation Component. The tools also had better reliability when used in Region A. The reliability of the LNOCAC tool for assessing provinces and cities as well as the Nutrition Situation Component of the tool for assessing barangays in cities were more prominent than in other tools. Nevertheless, results indicate that the reliability of the tools needs improvement particularly for evaluating lower levels of LGUs and in dimensions where the MET members had low percentages of IRR. Moreover, the reliability of the LNOCAC tools was observed when used by different groups, although inconsistent results were found when used to evaluate NPP in the provinces. The tools also generated lower performance scores for the LGUs compared to the existing M&E system, particularly at the barangay levels.

The general reliability of the LNOCAC tools observed may be attributed to the training workshop conducted among the MET members. According to Sattler *et al.* (2015), training can improve the IRR of raters as it leads to a common understanding of the definitions and meaning of

the rating scale. However, there is still a need to improve the reliability of the tools in producing consistent results for evaluating local NPP among MET members as low IRR was observed in their performance scores for the LGUs in some dimensions, which may be due to inconsistent implementation of a rating system (Lange 2011).

To improve the reliability of the tools, the IRR of the MET members must be improved. Based on the literature, this can be achieved through tool revisions and repeated instructions (Blick *et al.* 2018). Hence, further refinement and emphasis during the training activities of the tool dimensions in which the MET members had low percentages of acceptable IRR are warranted to reduce frequent rater errors and achieve a desirable level of reliability of the tools. Moreover, the reliability of the tools can also be secured by providing a more elaborate description of the dimensions used. Further, the evaluators may be given or shown an actual LGU scenario for each rating category in each dimension during the training for them to have a better knowledge base of the tools.

Additionally, the construction of the manual of operations and procedures, random testing for the percentage of agreement among the end-users, written and verbal communication options for end-users to address questions and problems, and considering fatigue relative to the time of use can enhance the IRR of the tool

Field trial of the LGNMES in the Philippines

Table 4 . Comparison of the evaluation scores using the LGNMES tool and the MELLPI

	Mean (Standard deviation)			p-value	
	MET	PT	MELLPI	MET vs MELLPI	PT vs MELLPI
Region A					
Barangay	51.78 (17.54)	42.04 (16.51)	88.28 (10.01)	<0.0001*	<0.0001*
Municipality	60.61 (22.79)	65.80 (20.85)	91.87 (7.18)	0.020*	0.051
City	44.37	48.82	87.87	-	-
Province	55.35 (12.32)	38.19 (16.85)	87.25 (8.55)	0.081	0.028*
Region B					
Barangay	61.35 (17.08)	51.03 (21.42)	91.38 (21.61)	0.002*	0.006*
Municipality	72.02 (15.92)	60.95 (17.15)	86.36 (19.21)	0.108	0.052
City	58.40 (14.53)	44.78 (17.12)	87.99 (3.54)	0.034*	0.011*
Province	55.54 (3.83)	54.72 (3.95)	90.98 (2.29)	0.077	0.077

*Significant at 5% level; MET: Monitoring and Evaluation Team; PT: Project Team; LGNMES: Local Government Nutrition Monitoring and Evaluation System; MELLPI: Monitoring and Evaluation of Local Level Plan Implementation

end-users (Burns 2014). Perhaps, developing a guideline for the MET members in using the LNOCAC tools to clarify the definition of its dimensions would improve the reliability of the tools. Periodic assessment of the IRR of the evaluators in using the tools and setting a limit for M&E sessions to avoid fatigue may also help reduce disagreement among evaluators and increase the reliability of the tools.

Using reliable M&E tools is essential to assess the effectiveness of nutrition programs, identify and address problems in program implementation, and disseminate data for public health actions to improve overall health (Jefferds & Flores-Ayala 2016). Other countries like Bangladesh were able to increase the stakeholders' commitment to community-based nutrition project and their understanding of its progress and evaluation activities by adopting a collaborative M&E system (Kang *et al.* 2021) that is supposed to result in a stronger evaluation design, enhanced data collection and analysis, and M&E data that stakeholders understand and use (O'Sullivan 2012). Moreover, a study showed that strengthening the M&E system could improve the performance of health-related projects (Micah & Luketero 2017).

However, it is important to note that the new system can generate significantly lower performance scores, especially at the barangay levels. Hence, caution should be taken when comparing the annual performance scores of the

LGUs that were based on different M&E systems. Nevertheless, its adoption would allow the assessment of the LGUs' performance in relation to the quality standards and evidence-based measures. It will also promote joint discussion among MET members and the LGUs assessed for learning and action plans for nutrition. Moreover, the system is consistent with the Philippine Plan of Action for Nutrition's strategic thrusts and can generate information that would be useful for planning programs for nutritionally at-risk pregnant women and young children.

The study provides evidence on the reliability of the proposed M&E tools in efforts to support the NNC's plan for updating the current M&E system. It is also supplemented with qualitative data regarding the insights from the target end-users of the tools which are deemed useful for future enhancement, updating, and nationwide implementation of the new system. However, the field trial of the tools was only limited to two regions in this study. Other concerns and issues regarding the system may still emerge particularly in LGUs with different organizational settings and M&E practices to the areas included in the study.

CONCLUSION

The MET members generally had a positive perception of the proposed M&E system for the evaluation of local NPP, although they

also raised critical issues and concerns that must be considered to minimize possible problems in the future. The reliability of the LNOCAC tools was also observed in general, but various steps are still warranted for improvement. Moreover, the adaption of the proposed M&E system may have implications for the performance scores of the LGUs.

ACKNOWLEDGEMENT

This research was funded by the NNC Central Office. The authors would like to thank the NNC Regional Offices for helping in the facilitation of the study as well as the MET members and the LGUs who participated in this study.

AUTHOR DISCLOSURES

The authors declare that they have no competing interests.

REFERENCES

- Blick B, Nakabugo S, Garabedian LF, Seru M, Trap B. 2018. Evaluating interrater reliability of indicators to assess performance of medicines management in health facilities in Uganda. *J Pharm Policy Pract* 11(1):1–12. <https://doi.org/10.1186/s40545-018-0137-y>
- Burns MK. 2014. How to establish interrater reliability. *Nursing* 2020 44(10):56–58. <https://doi.org/10.1097/01.NURSE.0000453705.41413.c6>
- Drummond KE, Murphy-Reyes A. 2018. *Nutrition Research: Concepts & Applications*. 1st Edition. Massachusetts (USA): Jones & Bartlett Learning.
- Gawe HC. 2015. Updated Nutrition Monitoring and Evaluation Protocol – Local Government NME Guidelines. Presented at the 5th National Conference of the District/City Nutrition Program Coordinators, 3–4th of November. Metro Manila, Philippines.
- George HA, Davis SL, Mitchell CF, Moyer NM, Toner CG. 2013. Abstraction of core measure data: Creating a process for interrater reliability. *J Nurs Care Qual* 28(1):68–75. <https://doi.org/10.1097/NCQ.0b013e31826ee49e>
- Kang Y, Cho M, Rahman MM, Cho Y, Han S, Dutta ML. 2021. Design of a collaborative monitoring and evaluation system for a community-based nutrition project in rural Bangladesh. *Eval Program Plann*, 84:101892. <https://doi.org/10.1016/j.evalprogplan.2020.101892>
- Jefferds MED, Flores-Ayala R. 2016. Chapter 5.6 What gets measured gets done: How nutrition monitoring, impact evaluation, and surveillance can support program improvement and policy development. In Eggersdorfer M, Kraemer K, Cordaro JB, Fanzo J, Gibney M, Kennedy E, Labrique A, Steffen J. *Good Nutrition: Perspectives for the 21st Century*. Basel, Karger (301–311). <https://doi.org/10.1159/000452394>
- Lange RT. 2011. Inter-Rater Reliability in *Encyclopedia of Clinical Neuropsychology*. New York (USA): Springer New York.
- Micah NJ, Luketero SW. 2017. Monitoring and evaluation systems and performance of non-governmental based maternal health projects in Bungoma South sub-county, Kenya. *Eur Sci J* 13(23):11–38. <https://doi.org/10.19044/esj.2017.v13n23p11>
- O'Sullivan RG. 2012. Collaborative evaluation within a framework of stakeholder-oriented evaluation approaches. *Eval Program Plann* 35(4):518–522. <https://doi.org/10.1016/j.evalprogplan.2011.12.005>
- Piedra-Fernández JM, Ganoza-Guerrero GE. 2016. Field trials and community trials: Importance in public health. *Medwave* 16(11):e6797. <https://doi.org/10.5867/medwave.2016.11.6797>
- Sattler DN, McKnight PE, Naney L, Mathis R. 2015. Grant peer review: Improving interrater reliability with training. *Plos One* 10(6):e0130450. <https://doi.org/10.1371/journal.pone.0130450>
- Smith PG, Morrow RH, Ross DA. 2015. *Field Trials of Health Interventions: A toolbox*. 3 ed. Oxford (UK): Oxford University Press.
- Vidyarani A, Martianto D, Syarief H. 2021. Evaluation of food and nutrition security level at provincial level based on outcome indicators in Indonesia. *J Gizi Pangan* 16(1):1–10. <https://doi.org/10.25182/jgp.2021.16.1.1-10>
- Wandner SA. 2017. *Lessons Learned from Public Workforce Program Experiments*. Kalamazoo (USA): W.E. Upjohn Institute for Employment Research.