JOURNAL OF INTERDISCIPLINARY PERSPECTIVES
PRINT ISSN: 2984-8288 | ONLINE ISSN: 2984-8385
https://doi.org/10.69569/jip

TWR Publishing

Research Article

"Are We Doing It Right?": The Dynamic CIPP Governance Model Evaluation Tool for Program Effectiveness

Gereo A. Patrimonio

Author Information:

Graduate Programs, Silliman University, Dumaguete City, Philippines

Corresponding Author:

Gereo A. Patrimonio
Email: Patrimonio.gereo@gmail.com

Article History:

Date received: September 21, 2026
Date revised: October 29, 2026
Date accepted: November 11, 2026

Recommended citation:

Patrimonio, G. (2026). "Are we doing it right?": The Dynamic CIPP Governance Model Evaluation Tool for program effectiveness. *Journal of Interdisciplinary Perspectives, 4*(12), 179-195. https://doi.org/10.69569/jip.2026.477

Abstract. Local government units require context-sensitive evaluation tools that incorporate beneficiaries' experiences into assessments of local health initiatives. However, the reviewed literature revealed a gap in culturally and linguistically adapted instruments for assessing beneficiaries' perceptions across the Context, Input, Process, and Product dimensions in Philippine local-government settings. This study developed and evaluated the Dynamic CIPP Governance Model Evaluation Tool using a sequential, two-phase design. Phase 1 involved participatory cultural and linguistic adaptation, pretesting, internal-consistency analysis, and exploratory factor analysis with 125 beneficiaries, parents or guardians of child beneficiaries, and local program personnel. Findings informed the refinement of selected items. In Phase 2, 701 respondents from two local health initiatives completed the revised 28-item instrument, comprising parents or guardians of child beneficiaries in one initiative and direct beneficiaries in the other. Its hypothesized correlated four-factor structure was evaluated through confirmatory factor analysis. Cronbach's alpha and McDonald's omega ranged from 0.88 to 0.90 across the four dimensions. The model demonstrated good fit, $\chi^2(344) = 360.29$, $p = .262$, comparative fit index = 0.998, Tucker-Lewis index = 0.998, root mean square error of approximation = 0.01, and standardized root mean square residual = 0.02. Standardized loadings ranged from 0.66 to 0.79. Average variance extracted ranged from 0.50 to 0.56, composite reliability from 0.88 to 0.90, and heterotrait-monotrait ratios from .43 to .52. The results supported the internal consistency, factorial validity, convergent validity, composite reliability, and discriminant validity of the four dimensions. The tool therefore provides local government units with a structured, beneficiary-oriented means of complementing administrative and outcome evidence when evaluating local health initiatives.

Keywords: *Confirmatory factor analysis; Dynamic CIPP Governance Model; Instrument development; Local government evaluation; Scale validation.*

Program evaluation is essential for determining whether public initiatives address identified needs, use resources appropriately, and achieve intended outcomes. Within local governance, evaluation enables decision-makers to identify implementation gaps, refine interventions, and enhance accountability through systematic evidence (Rossi et al., 2004; National Economic and Development Authority [NEDA], 2020; U.S. Government Accountability Office [GAO], 2021). This evaluative function is particularly critical in public health, where the relevance, accessibility, and quality of services directly influence community well-being. Strengthening the evaluation of local health programs also advances Sustainable Development Goal (SDG) 3 on Good Health and Well-being and SDG 16 on Peace, Justice, and Strong Institutions (United Nations General Assembly, 2015).

This work is licensed under a Creative Commons Attribution-Noncommercial 4.0 International License (CC BY-NC 4.0)

Although administrative records and outcome indicators are indispensable, they may not fully capture how intended beneficiaries receive and experience programs. Recent evaluation research in the Philippines underscores the importance of integrating beneficiaries' experiences into assessments of community-based health initiatives. For instance, a qualitative evaluation of health and nutrition programs by dela Luna *et al.* (2026) demonstrated that beneficiaries' accounts can reveal both perceived program benefits and implementation challenges related to resources, sustainability, and community participation. These findings underscore the significance of beneficiary feedback in understanding program relevance and identifying areas for improvement. Beneficiary perceptions regarding program relevance, accessibility, resource adequacy, communication, implementation, and benefits can influence participation and sustained engagement. Trust in public institutions and clear program communication further shape citizen responses to government initiatives (Rodrigues *et al.*, 2023). Conversely, misinformation and negative public narratives can distort perceptions and reduce participation, particularly in health-related programs (Ding *et al.*, 2023).

The Theory of Planned Behavior (TPB) provides a behavioral framework for understanding the significance of these perceptions in relation to stakeholder engagement. According to TPB, attitudes, subjective norms, and perceived behavioral control shape behavioral intentions and subsequent actions (Ajzen, 1991; Ajzen & Fishbein, 2005/2014). Empirical studies demonstrate that attitudes and social influences affect public responses to health interventions and other collective initiatives (Betsch *et al.,* 2015; Seddig *et al.,* 2022). Additional factors such as perceived credibility, social reinforcement, communication, and ease of participation further influence stakeholder engagement and ownership (Phang & Ilham, 2023; Wandani *et al.*, 2024). In local health initiatives, this behavioral perspective acknowledges that beneficiaries' participation and feedback are shaped by social and experiential contexts. Perception-based evidence should not replace administrative or outcome data; rather, it can complement these sources by clarifying how beneficiaries interpret and experience program delivery and by identifying concerns that conventional performance indicators may overlook (Prasetyo *et al.,* 2020).

While beneficiary perceptions provide valuable evidence regarding program experiences, evaluating effectiveness also requires determining whether initiatives achieve their intended objectives. Effectiveness refers to the attainment of intended objectives and meaningful results, whereas efficiency concerns the use of resources to produce those results (McCormick, 1981). Drucker (2006) emphasized that effectiveness depends on purposeful action and the disciplined prioritization of activities that most directly advance organizational goals. Consequently, evaluating public programs requires more than documenting completed activities or outputs. It necessitates examining whether needs were accurately identified, resources were sufficient, implementation was properly executed, and intended benefits were realized. The Context, Input, Process, and Product (CIPP) model provides a systematic framework for analyzing these interconnected aspects of a program across planning, implementation, and outcomes (Stufflebeam, 2003; G. Zhang *et al.*, 2011). Its application in educational, institutional, and governance contexts demonstrates its adaptability for both formative improvement and summative evaluation (Sundari, 2019; Wahjusaputri & Irawan, 2019; Sankaran & Saad, 2022).

In the Philippines, the National Evaluation Policy Framework (NEPF) institutionalizes evaluation as a mechanism for evidence-based governance (NEDA, 2020). However, local government units continue to encounter challenges such as limited technical capacity, political and administrative influences, fragmented information systems, inadequate baseline data, and limited access to assessment instruments suited to local contexts (Balgos, 2023; Corpus, 2022). These challenges are particularly consequential for local health initiatives, whose effectiveness depends on institutional performance, service accessibility, cultural relevance, public trust, and community participation. Persistent poverty further underscores the need for responsive and inclusive health services, as many Filipino households continue to identify themselves as poor (Social Weather Stations [SWS], 2025).

The literature reviewed for this study did not identify a validated, culturally and linguistically adapted instrument specifically designed to assess beneficiary perceptions of local health initiatives across the four CIPP dimensions within Philippine local-government settings. Although existing frameworks address program effectiveness, implementation, and behavioral engagement, these perspectives have not been sufficiently translated into a locally appropriate instrument that systematically captures beneficiaries' experiences of program relevance, resource adequacy, implementation, and perceived outcomes. In the absence of such an instrument, local governments may have limited means of incorporating structured beneficiary feedback into program evaluation and evidence-based decision-making.

Philippine researchers have applied the CIPP framework in various institutional and program settings. For example, Acenas (2023) used CIPP to evaluate an undergraduate research course, while Amoto *et al*. (2024) applied the framework to an institutional action research initiative. In the healthcare sector, Luib *et al.* (2022) evaluated patient medication counseling services using CIPP, and Catubig (2026) examined school-based nutrition programs through a CIPP-based evaluation model. Collectively, these studies demonstrate the adaptability of CIPP across Philippine educational, institutional, and health-related contexts. However, their evaluation purposes, respondent populations, and measurement approaches differ from the specific focus of the Dynamic CIPP Governance Model (DCGM) on culturally and linguistically adapted assessment of beneficiaries' perceptions across the four CIPP dimensions in Philippine local-government health initiatives.

Building on these applications, the present study developed the DCGM Evaluation Tool to address the identified measurement need. Although originally developed for educational evaluation, Stufflebeam's CIPP framework has demonstrated applicability across institutional and program settings. The DCGM extends this cross-contextual application to Philippine local governance without altering the theoretical definitions of Context, Input, Process, and Product. Its primary contribution is the operationalization of the established CIPP dimensions into a culturally and linguistically adapted instrument for systematically assessing beneficiaries' perceptions of local health initiatives. The TPB and principles of organizational effectiveness offer complementary perspectives on beneficiary engagement and program performance, respectively. The conceptual framework and operationalization of the four dimensions are detailed in the Methodology section.

The study aimed to develop and validate the DCGM Evaluation Tool for assessing stakeholders' perceptions of local health initiatives. Specifically, the objectives were to establish the instrument's linguistic clarity and cultural appropriateness through participatory review and pretesting, examine its preliminary factor structure and internal consistency during pilot administration, and evaluate the hypothesized correlated four-factor measurement model using an independent full-scale sample. The study also assessed the internal consistency, factorial validity, convergent validity, composite reliability, and discriminant validity of scores generated by the four dimensions. The resulting instrument is intended to provide local government units with a structured beneficiary-feedback mechanism that complements administrative and outcome evidence and supports program evaluation and improvement.

Methodology

Research Design and Validation Strategy

A sequential, two-phase approach was used for instrument development and validation. Phase 1 involved an exploratory evaluation of the initial 28-item DCGM Evaluation Tool to identify areas requiring refinement. Phase 2 consisted of a confirmatory evaluation of the revised instrument using an independent sample.

During Phase 1, item clarity, internal consistency, and preliminary factor structure were assessed to inform instrument refinement. In Phase 2, internal consistency was re-evaluated, and the hypothesized correlated four-factor measurement model was tested using confirmatory factor analysis (CFA). Convergent validity, composite reliability, and discriminant validity were also examined. The use of separate samples for the exploratory and confirmatory phases reduced the risk of evaluating the hypothesized measurement structure with the same data used for instrument refinement. Comparable contemporary scale-development studies have distinguished exploratory refinement from confirmatory evaluation through nonoverlapping samples or partitioned datasets (Metehan *et al.*, 2025; Khoury & Vergara, 2026).

Conceptualization and Development of the DCGM Evaluation Tool

The DCGM was developed to address the operational and contextual gap established in the Introduction by providing a culturally and linguistically adapted instrument for systematically assessing stakeholder perceptions of local health initiatives across the four CIPP dimensions. Addressing this gap required more than adopting an existing evaluation model. The framework needed to specify what should be evaluated, explain why perception-based evidence is consequential, and establish how the resulting evidence could support program decisions and improvement.

Accordingly, the DCGM integrated three distinct but equally necessary functions. The CIPP model provided the evaluative architecture by specifying the domains of Context, Input, Process, and Product (Stufflebeam, 2003; G.

Zhang *et al.*, 2011). The TPB provided the behavioral explanation for the significance of stakeholder perceptions and engagement (Ajzen, 1991). Principles of effectiveness provided the action-oriented logic for translating evaluation evidence into decisions, implementation adjustments, and continuing improvement (Drucker, 2006; McCormick, 1981). Figure 1 presents these functions within a continuing evaluation cycle. Together, they translate the study's central question, *Are we doing it right?*, into four operational concerns: whether the right community needs are being addressed, whether appropriate resources and institutional support are being provided, whether the initiative is being implemented properly, and whether meaningful and sustainable results are being achieved.

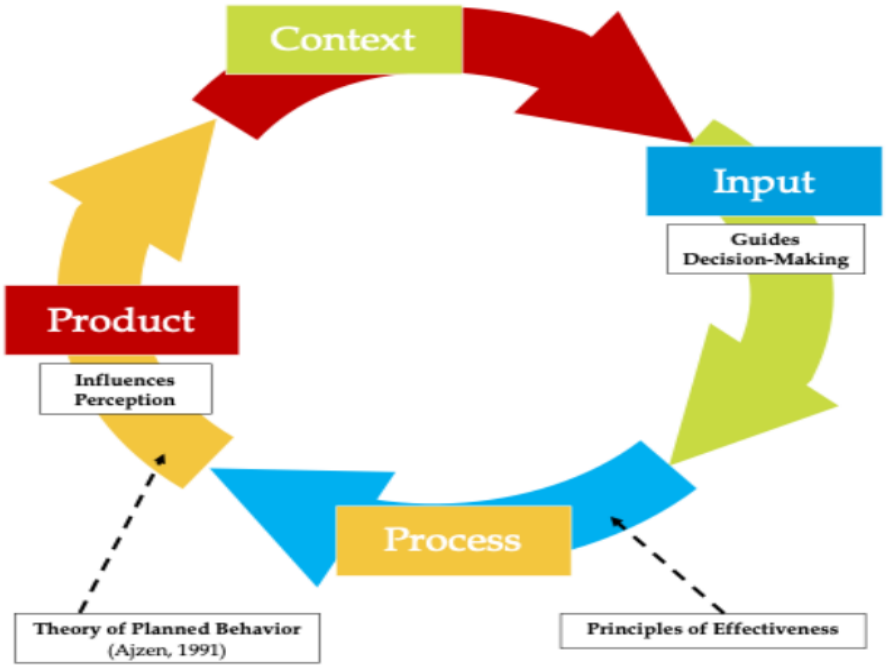


Figure 1. *The Dynamic CIPP Governance Model (DCGM)*

Context establishes the conditions to which an initiative should respond through indicators concerning community health needs, priorities, cultural appropriateness, participation in planning, and program awareness. Input examines whether the personnel, facilities, funding, materials, training, and community support required to address those conditions are available. The statement "guides decision-making" attached to Input signifies the point at which evidence is translated into choices about program feasibility, design, priorities, institutional support, and resource allocation. Although evidence from all four dimensions can inform program decisions, Input particularly connects the needs identified through Context with the means selected to address them.

These decisions are translated into action through Process, whose indicators address implementation, timeliness, accessibility, communication, coordination, community engagement, monitoring, and adaptability. The dashed connection from the principles of effectiveness indicates that implementation should not be judged merely by whether activities were completed but also by whether they addressed the right needs and advanced intended results. Effectiveness-oriented management emphasizes identifying what needs to be done, directing effort toward desired contributions, and reviewing results against expectations to guide subsequent action (Drucker, 2006). Effectiveness is similarly assessed in relation to clearly stated objectives and the acceptability of services to the people served (McCormick, 1981). These principles give the Process dimension its action-oriented function by connecting program decisions with purposeful implementation and corrective adjustment.

Table 1. *The 28-Item DCGM Evaluation Tool*

Construct	Item Descriptor	Measurement Item
Context	Q01-Addressing Critical Health Issues	I find that the health program addresses the most critical health issues in our community.
	Q02-Needs Assessment	I find that the health program is designed based on a thorough assessment of our community's needs.
	Q03-Alignment with Community Priorities	I find that the health program aligns well with the priorities of our community.
	Q04-Cultural Appropriateness	I find that the health program is culturally appropriate for our community.
	Q05-Community Involvement in Planning	I find that community members were involved in the planning of the health program.
	Q06-Program Awareness	I find that the community is well-informed about the health program.
	Q07-Addressing Specific Health Concerns	I find that the health program effectively addresses specific health concerns in our community.
Input	Q08-Adequacy of Resources	I find that the resources allocated for the health program are adequate.
	Q09-Quality of Personnel	I find that the personnel involved in the health program are highly qualified.
	Q10-Sufficiency & Accessibility of Facilities	I find that the facilities available for the health program are sufficient and accessible.
	Q11-Financial Support	I find that the financial support for the health program is sufficient.
	Q12-Quality of Materials & Equipment	I find that the materials and equipment provided for the health program are of good quality.
	Q13-Training & Development Opportunities	I find that the staff involved in the health program receive adequate training and development opportunities.
	Q14-Community Contributions	I find that the community contributes resources like time and effort to support the health program.
Process	Q15-Effectiveness of Implementation	I find that the health program is being implemented effectively.
	Q16-Timeliness of Services	I find that the services provided by the health program are delivered in a timely manner.
	Q17-Accessibility of Services	I find that the services provided by the health program are easily accessible to all community members.
	Q18-Communication & Coordination	I find that there is effective communication and coordination among the staff involved in the health program.
	Q19-Community Engagement	I find that the community is actively engaged in the implementation of the health program.
	Q20-Monitoring & Evaluation	I find that the health program is regularly monitored and evaluated to ensure quality.
	Q21-Adaptability to Changing Needs	I find that the health program is adaptable to changing community needs and circumstances.
Product	Q22-Positive Outcomes in Community Health	I find that the health program has led to positive outcomes in the community's health.
	Q23-Satisfaction with Services	I find that community members are satisfied with the services provided by the health program.
	Q24-Achievement of Goals	I find that the health program has achieved its intended goals.
	Q25-Impact on Health Behaviors	I find that the health program has positively impacted health behaviors in the community.
	Q26-Sustainability of Benefits	I find that the benefits of the health program are sustainable over the long term.
	Q27-Improvement in Overall Community Health	I find that the overall health of the community has improved as a result of the health program.
	Q28-Unintended Positive Outcomes	I find that the health program has resulted in positive outcomes that were not initially anticipated.

Note. The culturally and linguistically adapted Camotes Visayan version of the DCGM Evaluation Tool is available upon reasonable request.

Product examines perceived outcomes, satisfaction, goal attainment, behavioral effects, sustainability, and broader community-health benefits. The dashed connection from TPB to Product represents the conceptual importance of perceptions and engagement in understanding how an initiative is received and experienced. Stakeholders' interpretations of program relevance, accessibility, institutional support, and expected benefits may influence their engagement with the services provided. The TPB constructs were not measured as separate dimensions in the present instrument; instead, the theory explains why perception-based evidence is consequential within the framework. The statement "influences perception" under Product completes the feedback relationship. Experiences and interpretations of program outcomes may reshape subsequent perceptions, affect later acceptance and engagement, and provide evidence for future decisions. Product is

therefore not the terminal point of evaluation because its evidence can prompt a renewed assessment of Context, revised Input decisions, and adjustments to Process.

This interaction establishes the dynamic character of the DCGM. CIPP identifies what is evaluated; TPB explains why stakeholders' perceptions are consequential; effectiveness principles explain how the resulting evidence can guide action; "guides decision-making" connects identified needs and available means with program choices; and "influences perception" returns experienced outcomes to the continuing evaluation cycle. The framework was operationalized through a deductive item-development process using seven indicators for each CIPP dimension, resulting in a 28-item instrument: Context, Q01–Q07; Input, Q08–Q14; Process, Q15–Q21; and Product, Q22–Q28. The four CIPP dimensions constituted the measurement structure evaluated in this study, while TPB and effectiveness principles provided the explanatory and action-oriented functions through which the resulting scores could be interpreted and used.

Table 1 presents the item descriptors and corresponding English-language measurement statements of the DCGM Evaluation Tool. The instrument employed an intentionally asymmetric five-point agreement scale: 1 = Disagree (no agreement), 2 = Slightly Agree (1%–25% agreement), 3 = Moderately Agree (26%–50% agreement), 4 = Strongly Agree (51%–75% agreement), and 5 = Extremely Agree (76%–100% agreement). Unlike a balanced scale that offers multiple degrees of disagreement and a neutral midpoint, this structure provided a single category for nonagreement and four categories to differentiate levels of affirmative assessment. The percentage intervals served as interpretive anchors rather than as independently measured proportions. This response structure was developed to address the study's central evaluative question: *Are we doing it right?* In this context, effectiveness means achieving clearly stated objectives, including outcomes that address the target population's perceived needs (McCormick, 1981). Effectiveness-oriented management similarly emphasizes identifying necessary actions, defining desired results, and comparing actual outcomes with expectations to guide future decisions (Drucker, 2006). The graded agreement categories allowed respondents to indicate the extent to which they perceived that the local health initiative addressed community needs, mobilized appropriate resources, implemented services effectively, and produced intended outcomes. Higher scores indicated more favorable evaluations of each program dimension and provided nuanced evidence to inform program review and improvement.

Participatory Content Review, Cultural and Linguistic Adaptation, and Pretesting

The initial measurement items were developed in English and then translated into the Camotes Islands variety of Visayan. Personnel from the Office of the Municipal Administrator and the Office of the Mayor helped prepare the initial translation. The translation process emphasized preserving each item's intended meaning rather than producing literal, word-for-word equivalents because lexical equivalents may not convey the same meaning across languages and cultural contexts (Altarriba, 1992).

The first round of participatory review took place during a regular monthly meeting of the Barangay Daycare Workers (BDWs). The initial Visayan version was presented to the participants for review of its wording, clarity, comprehensibility, dialect usage, and suitability for community respondents. BDWs were selected as community-based reviewers because their regular engagement with residents at the barangay level provided practical familiarity with the language and expressions used by the intended respondents. Their feedback was used to refine words and statements that appeared unnatural, unclear, or inconsistent with local usage. This procedure recognized that language conveys culturally situated meanings that direct translation may not fully preserve (Rivera & Altarriba, 2002).

The revised Visayan instrument was pretested during Phase 1 with beneficiaries, parents or guardians of child beneficiaries, and personnel involved in the two local health initiatives. In addition to completing the instrument, participating barangay and municipal personnel provided field-based feedback on dialect usage, translation clarity, and item comprehension. This feedback and evidence from the Phase 1 internal consistency and exploratory factor analyses (EFA) were considered when determining whether specific items required further review. Comparable scale-development research has used linguistic review and participant pretesting to determine whether draft items are clear and understandable before broader psychometric evaluation (Çelebi *et al.*, 2025).

A second round of participatory review was conducted after the Phase 1 exploratory analysis. Items with comparatively weak primary loadings were presented during the respective regular monthly meetings of the BDWs, Barangay Health Workers (BHWs), and Barangay Nutrition Scholars (BNSs). The review focused on whether the wording clearly conveyed the intended meanings in the local variety of Visayan. Drawing on their direct engagement with community members and familiarity with local language use, the BDWs, BHWs, and BNSs contributed to refining the wording of the flagged items. Item retention was based on linguistic clarity, contextual appropriateness, conceptual relevance, and the need to maintain adequate coverage of the intended DCGM dimensions. The refined 28-item Visayan instrument was then prepared for Phase 2 administration.

Participants and Data-Collection Procedures

The study was conducted in Tudela, Cebu, using respondent groups aligned with the distinct purposes of the two validation phases. Phase 1 employed a heterogeneous developmental sample comprising beneficiaries, beneficiary representatives, and personnel familiar with the two local health initiatives and the communities they served. This composition supported preliminary statistical assessment and field-based review of item clarity and contextual appropriateness. Phase 2 focused on the intended respondents of the instrument, comprising beneficiaries and parents or guardians representing child beneficiaries of the same initiatives. Phase 1 data were collected in January 2025, while Phase 2 data were collected from April to May 2025.

The Phase 1 sample consisted of 125 respondents from three groups: beneficiaries and beneficiary representatives, barangay workers, and municipal personnel. Beneficiaries and their representatives were recruited through snowball sampling at the Rural Health Unit (RHU) and barangay health centers. Parents or guardians completed the pilot instrument as adult representatives of child beneficiaries of Local Initiative A, while beneficiaries of Local Initiative B completed the instrument themselves. Barangay workers were invited to participate due to their direct engagement with beneficiaries and familiarity with community conditions and local language use. Of the 65 barangay workers invited, 51 voluntarily participated. Municipal personnel were invited based on their administrative involvement and knowledge of the initiatives. Of the 16 municipal personnel invited, nine voluntarily participated. Nonparticipation among barangay and municipal personnel was due to absence or a decision not to participate, and no individual was compelled to complete the instrument. Barangay and municipal personnel were also invited to provide feedback on item clarity and comprehensibility. Their familiarity with program implementation, intended respondents, and local language use made their participation relevant to the developmental purpose of Phase 1. The quantitative responses and field-based feedback were used to evaluate and refine the instrument. Phase 1 data were not included in the subsequent confirmatory evaluation. Table 2 presents the composition of the Phase 1 sample.

Table 2. *Respondent Composition in the Pilot Testing*

Respondent Group	Sample Size (*n*)	Percentage (%)
Beneficiaries and beneficiary representatives:	65	52.00%
Parents/guardians of Local Initiative A child beneficiaries	32	49.23%
Local Initiative B beneficiaries	33	50.77%
Barangay Workforce:	51	40.80%
Barangay Health Workers (BHW)	25	49.02%
Barangay Daycare Workers (BDW)	15	29.41%
Barangay Nutrition Scholars (BNS)	11	21.57%
Municipal Government Workforce:	9	7.20%
Mun. Social Welfare & Development Office (MSWDO)	4	44.44%
Rural Health Unit (RHU)	5	55.56%
Total	125	100.00%

Note. Percentages for the three principal respondent groups were calculated using the total Phase 1 sample (N = 125). Percentages for the subgroups were calculated within their respective principal respondent groups. Percentages may not total exactly 100 due to rounding.

Phase 2 sought complete enumeration using the cumulative official beneficiary lists of the two local health initiatives. The list for Local Initiative A comprised 416 child beneficiaries registered from 2020 to 2025, while the list for Local Initiative B comprised 383 beneficiaries registered from 2021 to 2025. For Local Initiative A, an adult parent or guardian, either a father, mother, or grandparent, was invited to complete the questionnaire as the representative of each registered child beneficiary. Registered beneficiaries of Local Initiative B were invited to complete the questionnaire themselves. The sampling frame therefore comprised 799 beneficiary records accumulated across the respective implementation periods rather than records from a single program year. Data

collection yielded 701 valid responses for the final Phase 2 analytic sample. These responses represented assessments from beneficiaries and beneficiary representatives and were analyzed independently to reassess internal consistency and evaluate the hypothesized four-factor measurement structure. Initiative-specific response counts and response rates are presented in Table 4. In both phases, the culturally and linguistically adapted Visayan questionnaire was administered in printed, paper-based form and completed independently by the respondents. Upon submission, survey administrators checked each questionnaire for unanswered items. Respondents were given the opportunity to complete any omissions before final submission. All returned questionnaires were complete and included in the respective analyses. The study therefore obtained 125 valid questionnaires in Phase 1 and 701 valid questionnaires in Phase 2, with no returned questionnaires excluded because of missing responses.

Data Analysis

Data from the two validation phases were analyzed separately. Phase 1 analyses were conducted using IBM SPSS Statistics Version 30. Phase 2 analyses were conducted using R Version 4.4.3 through RStudio Version 2026.09.0 Build 174. Phase 2 data management, reliability assessment, confirmatory factor analysis, and validity assessment used the R packages *psych* Version 2.6.1, *lavaan* Version 0.6-19, *semTools* Version 0.5-6, and *dplyr* Version 1.1.4. Statistical significance was set at the .05 level unless otherwise specified.

During Phase 1, the preliminary internal consistency of the four DCGM dimensions and the complete instrument was assessed using Cronbach's alpha. A coefficient of 0.70 was used as the minimum acceptable criterion for this developmental phase (Nunnally, 1978). The suitability of the pilot data for EFA was evaluated using the Kaiser-Meyer-Olkin measure of sampling adequacy and Bartlett's test of sphericity. Principal axis factoring with Varimax rotation and Kaiser normalization was used to examine the instrument's preliminary internal structure. Factors with eigenvalues greater than 1.00 were initially retained, and 0.45 was used as the primary criterion for interpreting factor loadings (Tabachnick & Fidell, 2014). Items with weak or cross-factor loadings were reviewed together with field-based feedback concerning translation, contextual relevance, and comprehensibility. Decisions to retain or revise an item considered both its statistical performance and its theoretical relevance to the intended DCGM dimension.

During Phase 2, means and standard deviations were calculated from the mean item scores for each DCGM dimension and the complete instrument. Internal consistency was assessed using Cronbach's alpha and McDonald's omega, with 0.70 used as the minimum acceptable criterion. The hypothesized measurement structure was evaluated through CFA using robust maximum likelihood estimation (MLR). Given the five response categories and the large Phase 2 sample, item scores were treated as approximately continuous. MLR was selected to provide robust standard errors and scaled test statistics under potential nonnormality. Context, Input, Process, and Product were specified as four correlated latent constructs, each measured by seven observed indicators. Model fit was evaluated using the robust chi-square statistic, comparative fit index, Tucker-Lewis index, root mean square error of approximation with its 90% confidence interval, and standardized root mean square residual (Kline, 2016).

Convergent validity was evaluated using standardized factor loadings and average variance extracted. Average variance extracted values of at least 0.50 were considered acceptable. Composite reliability was calculated from the standardized CFA estimates, with values of at least 0.70 considered acceptable (Hair *et al.,* 2022). Discriminant validity was assessed using the Fornell-Larcker criterion and the heterotrait-monotrait ratio of correlations. The Fornell-Larcker criterion was satisfied when the square root of each construct's average variance extracted exceeded its correlations with the other constructs. Heterotrait-monotrait values below .85 were interpreted as supporting discriminant validity (Fornell & Larcker, 1981; Henseler *et al.,* 2015).

Collectively, these analyses evaluated the reliability and construct validity of scores generated by the 28 items across the four CIPP dimensions specified in the DCGM.

Ethical Consideration

The Municipality of Tudela conducted this study as part of an officially approved program-evaluation and research initiative. The research activity proposal, approved on April 16, 2024, defined the study's objectives, methodology, and implementation procedures (Municipality of Tudela, 2024). Executive Order No. 19-B, Series

of 2024, issued on April 25, 2024, established and authorized an activity team to implement the study (Office of the Mayor of Tudela, 2024). The order appointed the researcher, then serving as Municipal Administrator, as the team convenor responsible for leading and coordinating the study and overseeing the development, implementation, and monitoring of the evaluation tool. Co-convenors, members, and auxiliary members were assigned technical, administrative, implementation, and data-collection responsibilities. During the study period, the municipality did not have a research ethics committee, and independent ethics approval was not obtained. The approved proposal and executive order provided institutional authorization and administrative accountability; however, these did not constitute independent research ethics clearance.

In both phases of the study, respondents were informed of the study's purpose, procedures, voluntary participation, and privacy and confidentiality protections before completing the questionnaire. In Phase 1, the researcher personally presented this information to each participating group and addressed any questions before distributing the questionnaires. In Phase 2, the local government used official beneficiary lists to identify and contact intended respondents. Barangay Day Care Workers (BDWs) contacted or visited the parents or guardians of child beneficiaries of Local Initiative A, while Barangay Health Workers (BHWs) contacted or visited beneficiaries of Local Initiative B. Before data collection, the researcher oriented questionnaire administrators on study procedures and their responsibilities regarding voluntary participation, respondent protection, anonymity, confidentiality, and compliance with Republic Act No. 10173, the Data Privacy Act of 2012. These safeguards were incorporated into the approved research protocol, which explicitly required questionnaire administrators to receive orientation on data privacy, anonymity, and confidentiality (Municipality of Tudela, 2024).

The first page of the questionnaire included an anonymity and confidentiality statement in the local variety of Visayan, informing respondents of the study's purpose, voluntary participation, protection of their responses, intended use of the collected information, and potential publication of findings. The statement also explained respondents' right to withdraw participation and request the removal of unprocessed data. Before questionnaire administration, designated administrators verbally explained the statement and guided respondents through its provisions to ensure they understood their rights and the measures established to protect their responses. Any questions or concerns were addressed before respondents proceeded with the questionnaire. Separate signed informed-consent forms were not collected to avoid obtaining unnecessary identifying information and to minimize concerns about confidentiality. Instead, respondents received information about the study through the written statement and its verbal explanation. Voluntary completion and submission of the questionnaire indicated willingness to participate. Respondents could decline participation or discontinue answering without penalty or adverse consequence. Because completed questionnaires were not linked to respondents' identities, individual responses could not be retrieved for withdrawal after anonymous submission.

Data collection and handling were guided by the principles of Republic Act No. 10173, emphasizing the protection of personal information, confidentiality, and restricting data processing to legitimate research purposes. Although official beneficiary lists were used to identify and contact intended respondents, the study did not link identifying information from these lists to completed questionnaires. Respondents were not required to provide names or other identifying details. During data processing, any identifying information voluntarily disclosed in questionnaire responses was removed to preserve anonymity. Upon submission, questionnaire administrators checked responses only for unanswered items and allowed respondents to complete any omissions before final acceptance. Access to completed questionnaires and encoded data was limited to authorized research personnel, and the collected information was used exclusively for statistical and research purposes. Responses were aggregated, and findings were reported only in aggregate form. Respondents were also informed that the findings could be presented at a research conference or published in a scientific journal without disclosing their identities.

Results and Discussion

The results are organized according to the study's two validation phases. Phase 1 reports the preliminary internal consistency and exploratory factor structure of the DCGM Evaluation Tool and explains how the pilot evidence informed item refinement. Phase 2 presents response coverage, descriptive statistics, internal consistency, factorial validity, convergent validity, composite reliability, and discriminant validity using an independent respondent sample. The findings are interpreted in relation to the instrument-development objectives and the functional logic of the DCGM, with connections to relevant literature and implications for the interpretation and practical use of the resulting scores.

Phase 1: Pilot Internal Consistency and Exploratory Factor Structure

Phase 1 evaluated the preliminary internal consistency and exploratory factor structure of the 28-item DCGM Evaluation Tool using responses from 125 participants. Reliability analysis indicated satisfactory internal consistency within each theoretically specified CIPP dimension and across the entire instrument, with all coefficients exceeding the developmental criterion of 0.70 (Nunnally, 1978). These findings supported retaining the item sets for further evaluation. EFA was then conducted to examine the empirical relationships among the items. The Kaiser-Meyer-Olkin measure and Bartlett's test of sphericity indicated that the pilot responses were suitable for factor analysis. Principal axis factoring with Varimax rotation and Kaiser normalization yielded a five-factor exploratory solution, indicating that the empirical grouping of items did not correspond directly to the four theoretically specified CIPP dimensions. Table 3 presents the rotated factor loadings. Using 0.45 as the primary interpretive criterion (Tabachnick & Fidell, 2014), four items exhibited comparatively weak primary loadings: Q05, Community Involvement in Planning (.40); Q06, Program Awareness (.41); Q12, Quality of Materials and Equipment (.41); and Q03, Alignment with Community Priorities (.42).

Table 3. *Exploratory Factor Analysis (EFA) Results for DCGM Items in the Pilot Study*

Item	Factor				
	1	2	3	4	5
Q20	**0.69**				
Q26	**0.64**				
Q15	**0.60**				
Q19	**0.59**				
Q22	**0.58**				
Q16	**0.58**	0.55			
Q27	**0.55**	0.47			
Q25	**0.51**				
Q05	**0.40**				
Q08		**0.70**			
Q11		**0.64**		0.41	
Q17		**0.57**			
Q21		**0.53**			
Q02			**0.60**		
Q04			**0.57**		
Q23			**0.55**		
Q28			**0.50**		
Q03			**0.42**		
Q06			**0.41**		
Q10		0.41		**0.59**	
Q12			**0.41**		
Q18				**0.52**	
Q01	0.43			**0.50**	
Q14					**0.67**
Q09					**0.67**
Q07	0.47				**0.51**
Q24					**0.49**
Q13				0.40	**0.46**

Note. Item codes and their corresponding descriptors and measurement statements are presented in Table 1. Extraction method: principal axis factoring. Rotation method: Varimax with Kaiser normalization. Coefficients below 0.40 were suppressed. Bold values indicate the largest loading for each item. Rotation converged in 10 iterations.

These four items represented important aspects of the DCGM's intended measurement scope, particularly community participation, program awareness, resource adequacy, and responsiveness to community priorities. Removing them solely on the basis of their pilot loadings could have reduced the instrument's coverage of relevant program conditions and resources. Consequently, the exploratory findings informed the second participatory review described in Section 2.3, during which the flagged items underwent linguistic and contextual refinement while retaining their intended conceptual meanings. This approach was consistent with the study's emphasis on preserving conceptual meaning during linguistic adaptation rather than relying exclusively on literal word equivalence (Altarriba, 1992; Rivera & Altarriba, 2002).

Phase 1 provided preliminary evidence of internal consistency and identified specific areas requiring linguistic and contextual refinement. All 28 items and their intended CIPP dimensions were retained, with improvements to the clarity of selected measurement statements. The refined instrument was subsequently administered to an independent Phase 2 sample for confirmatory evaluation of its hypothesized four-factor measurement structure.

Phase 2: Administration, Descriptive Statistics, and Internal Consistency

In Phase 2, valid responses were collected from beneficiaries and their representatives of the two local health initiatives, using the municipality's official beneficiary records as the sampling frame. For Local Initiative A, parents or guardians responded on behalf of child beneficiaries, while beneficiaries of Local Initiative B completed the instrument directly. The comparable response rates across both initiatives indicated similar participation coverage. The resulting sample was independent of the Phase 1 pilot sample and provided the data for confirmatory validation of the DCGM Evaluation Tool. Table 4 presents the beneficiary counts, actual respondents, and corresponding response rates.

Table 4. *Summary of Beneficiaries and Respondents*

Program	Total Beneficiaries	Actual Respondents	Response Rate (%)
1. Local Initiative A	416	366	87.98
2. Local Initiative B	383	335	87.47
Total	799	701	87.73

Note. Beneficiary totals represent the cumulative official beneficiary lists for 2020–2025 for Local Initiative A and 2021–2025 for Local Initiative B; they do not represent enrollment in a single program year. For Local Initiative A, parents or guardians completed the questionnaire on behalf of the registered child beneficiaries, whereas the beneficiaries of Local Initiative B completed the questionnaire themselves.

Table 5 presents descriptive statistics and internal consistency estimates for the four DCGM dimensions and the complete instrument. The mean scores indicated generally affirmative respondent assessments across all dimensions, consistent with the agreement structure described in Section 2.2. However, the observed variability suggests that respondents differentiated their assessments rather than expressing uniformly favorable judgments, reflecting differences in their perceptions of the initiatives. Both Cronbach's alpha and McDonald's omega exceeded the minimum criterion of 0.70 (Nunnally, 1978), with closely corresponding estimates providing consistent evidence of strong internal consistency within each dimension.

Table 5. *Phase 2 Descriptive Statistics and Internal Consistency of the DCGM Tool (N = 701)*

DCGM dimension	No. of items	*M*	*SD*	Cronbach's α	McDonald's ω
Context	7	3.93	0.83	0.89	0.89
Input	7	3.96	0.81	0.88	0.88
Process	7	4.04	0.81	0.90	0.90
Product	7	4.10	0.79	0.89	0.89
Overall DCGM tool	**28**	**4.01**	**0.62**	**0.93**	**0.93**

Note. Mean and *SD* are based on mean item scores on the 1–5 response scale.

These findings support the internal consistency of scores generated by the four prespecified DCGM dimensions in the Phase 2 sample, providing a basis for examining community relevance, resources and institutional support, implementation, and perceived outcomes, as outlined in Section 2.2. Although the complete instrument also demonstrated strong internal consistency, this finding alone does not establish a unidimensional measurement structure or justify interpreting all items as a single overall DCGM score**.** The next stage of validation therefore examined whether responses to the refined instrument supported the hypothesized four-factor structure and the empirical distinctness of its dimensions through CFA and subsequent validity assessments**.**

Phase 2: Four-Factor Measurement Model

CFA was performed on the independent Phase 2 sample (N = 701) to assess the hypothesized correlated four-factor measurement structure of the DCGM Evaluation Tool. The model defined Context, Input, Process, and Product as four latent constructs, each measured by seven designated indicators. Table 6 presents the standardized factor loadings and error variances.

The four items identified for refinement during Phase 1 (Q03, Q05, Q06, and Q12) were of particular significance. After participatory linguistic and contextual refinement, all four items exhibited substantial standardized loadings on their designated constructs in the independent Phase 2 sample, as presented in Table 6. These results support the retention of these items and are consistent with the refinement process's objective of preserving their intended conceptual meanings while improving their linguistic and contextual appropriateness for the local respondent population.

Table 6. *Standardized Factor Loadings for the Four-Factor DCGM Measurement Model (N = 701)*

Construct	Item	λ	θ
Context	Q01	0.71	0.49
	Q02	0.73	0.47
	Q03	0.72	0.48
	Q04	0.76	0.43
	Q05	0.73	0.47
	Q06	0.74	0.46
	Q07	0.72	0.48
Input	Q08	0.68	0.54
	Q09	0.73	0.46
	Q10	0.74	0.45
	Q11	0.68	0.53
	Q12	0.73	0.47
	Q13	0.72	0.49
	Q14	0.68	0.54
Process	Q15	0.69	0.52
	Q16	0.75	0.44
	Q17	0.79	0.37
	Q18	0.73	0.47
	Q19	0.76	0.42
	Q20	0.74	0.46
	Q21	0.76	0.43
Product	Q22	0.66	0.56
	Q23	0.78	0.40
	Q24	0.75	0.44
	Q25	0.74	0.45
	Q26	0.72	0.48
	Q27	0.76	0.43
	Q28	0.72	0.48

Note. Item codes and their corresponding descriptors and measurement statements are presented in Table 1. λ = standardized factor loading; θ = standardized error variance. Estimates were obtained using robust maximum likelihood estimation. All factor loadings were statistically significant at $p < .001$.

The model demonstrated good fit to the observed data. Robust maximum likelihood estimation yielded a nonsignificant robust chi-square test, $\chi^2(344) = 360.29$, $p = .262$. Additional fit indices were robust CFI = 0.998, robust TLI = 0.998, robust RMSEA = 0.009, 90% CI [0.00, 0.02], and SRMR = 0.024. Collectively, these indices indicated close correspondence between the hypothesized four-factor structure and the observed data (Kline, 2016). The favorable model fit may also reflect, in part, the linguistic and contextual refinements undertaken during instrument development. By presenting measurement statements in culturally appropriate Visayan, the refinement process sought to facilitate respondents' comprehension of the intended concepts and their ability to relate these concepts to their experiences with the local health initiatives. This consideration is particularly relevant in community-based evaluations involving respondents with varying educational backgrounds. Although the analysis did not isolate the statistical contribution of linguistic adaptation, the findings were consistent with the study's emphasis on conceptual clarity and contextual appropriateness in instrument development.

All 28 items exhibited significant loadings on their respective constructs. Error variance estimates were within acceptable ranges, with no negative variances or other improper solutions. These findings provided item-level evidence consistent with the hypothesized measurement structure. The global fit and item-level results supported the proposed four-dimensional measurement structure of the DCGM Evaluation Tool. This structure provides a foundation for examining community relevance, resources and institutional support, implementation, and perceived outcomes as related but theoretically distinct aspects of program evaluation. The measurement properties of the four dimensions were further assessed through average variance extracted, composite reliability, and discriminant validity analyses.

Assessment of Convergent Validity and Composite Reliability

Convergent validity and composite reliability were evaluated using average variance extracted (AVE) and composite reliability (CR), respectively. Table 7 shows that all four DCGM dimensions met the minimum AVE threshold of 0.50 (Fornell & Larcker, 1981), indicating that their indicators shared sufficient variance to support convergent validity. Although the Input dimension had the lowest AVE, its unrounded value of 0.502 confirmed

that it exceeded the required criterion. The minor differences among the AVE estimates indicated similar levels of explained indicator variance across the four dimensions. Composite reliability estimates surpassed the minimum threshold of 0.70 for all dimensions, supporting the internal consistency of their respective indicators. The close correspondence with McDonald's omega coefficients reported in Table 5 provided complementary reliability evidence from the CFA-based and earlier reliability analyses.

Table 7. *Convergent Validity and Composite Reliability of the DCGM Dimensions (N = 701)*

DCGM Dimension	AVE	CR
Context	0.53	0.89
Input	0.50	0.88
Process	0.56	0.90
Product	0.54	0.89

Note. AVE = average variance extracted; CR = composite reliability. Values were calculated from the standardized CFA estimates.

Taken together, the AVE and CR results supported the convergent validity and composite reliability of the four prespecified DCGM dimensions. This within-dimension coherence is important for generating evidence regarding community relevance, resources and institutional support, implementation, and perceived outcomes**.** The integration of participatory linguistic refinement and subsequent psychometric assessment is consistent with Zhang and Aryadoust's (2022) emphasis on drawing upon multiple sources of evidence to establish questionnaire validity, rather than relying on statistical indicators alone. However, convergence within each dimension alone does not establish that the dimensions are empirically distinct. Subsequent analyses therefore examined discriminant validity to determine whether Context, Input, Process, and Product represented sufficiently distinguishable constructs.

Assessment of Discriminant Validity

Table 8. *Fornell-Larcker Assessment of Discriminant Validity for the DCGM Dimensions (N = 701)*

Construct	Context	Input	Process	Product
Context	**0.73**			
Input	.43	**0.71**		
Process	.50	.50	**0.75**	
Product	.47	.52	.50	**0.73**

Note. Bold diagonal values are the square roots of average variance extracted. Off-diagonal values are latent construct correlations. Blank upper-triangular cells indicate duplicate comparisons that were omitted.

Discriminant validity among Context, Input, Process, and Product was assessed using both the Fornell-Larcker criterion and the heterotrait-monotrait ratio of correlations (HTMT). While the Fornell-Larcker criterion is widely adopted, Henseler *et al.* (2015) demonstrated that it is less sensitive to specific discriminant validity concerns and recommended HTMT as a more robust alternative. Therefore, both approaches were employed to evaluate whether the four theoretically related DCGM dimensions could be empirically differentiated. The Fornell-Larcker criterion involves comparing the square root of each construct's average variance extracted (AVE) to its correlations with other constructs. Discriminant validity is established when the square root of AVE exceeds the corresponding interconstruct correlations (Fornell & Larcker, 1981). As shown in Table 8, all four dimensions met this criterion. The results indicated that each dimension shared greater variance with its designated indicators than with other dimensions, as determined by the Fornell-Larcker criterion. Although positive correlations were observed among the dimensions, these relationships align with the DCGM framework, which conceptualizes Context, Input, Process, and Product as interconnected yet distinct components of program evaluation.

To address the limitations of the Fornell-Larcker criterion in detecting discriminant validity issues, HTMT was also used to further examine discriminant validity. Henseler *et al.* (2015) recommended a conservative threshold of 0.85, although values below 0.90 may be acceptable for conceptually similar constructs. As indicated in Table 9, all HTMT ratios satisfied the more stringent threshold. The HTMT results supported the Fornell-Larcker findings, confirming that the four dimensions remained empirically distinguishable under a more sensitive assessment. The Input and Product pair demonstrated the strongest association in both analyses, reflecting their conceptual relationship while remaining empirically distinct.

Table 9. *HTMT Ratios for the DCGM Dimensions (N = 701)*

Construct	Context	Input	Process	Product
Context				
Input	.43			
Process	.49	.50		
Product	.47	.52	.51	

Note. HTMT = heterotrait-monotrait ratio of correlations. Blank diagonal cells indicate nonapplicable self-comparisons, while blank upper-triangular cells indicate duplicate comparisons that were omitted.

Collectively, the Fornell-Larcker and HTMT results demonstrated that Context, Input, Process, and Product were related yet empirically distinct within the Phase 2 sample. This distinction is essential to the functional logic of the DCGM. The relationships among the dimensions allow them to operate within a continuous evaluation cycle, while their empirical separation enables identification of whether perceived concerns primarily involve community relevance, resources and institutional support, implementation, or perceived outcomes. Repke and Felderer (2025) emphasized the importance of language and cultural context in survey measurement quality, offering a methodological perspective for interpreting these findings alongside the DCGM's participatory linguistic refinement process. Although the present study did not isolate the statistical contribution of linguistic adaptation, the discriminant validity results provide further evidence that the four intended constructs remained empirically distinguishable following refinement. Consequently, the dimensions offer differentiated evidence for program review and decision-making, rather than serving as interchangeable indicators of a single general perception.

The DCGM: Substance and Implication

The statistical results provide robust evidence supporting the validity of the DCGM Evaluation Tool. Highly favorable fit indices, such as CFI and TLI values of 0.998, were achieved using a theoretically specified four-factor model with substantial degrees of freedom (df = 344), rather than a saturated model. Furthermore, significant factor loadings, strong internal consistency, and satisfactory convergent and discriminant validity reinforce these findings. Collectively, the results demonstrate that the four dimensions measure related yet empirically distinct aspects of program effectiveness. This interpretation is consistent with Zhang and Aryadoust's (2022) assertion that questionnaire validity should be established through multiple sources of evidence rather than reliance on individual statistical indicators.

The instrument's development process establishes a methodological basis for interpreting these results. Pilot testing identified four items (Q03, Q05, Q06, and Q12) requiring further examination, which led to participatory linguistic and contextual refinement before administration to an independent sample. Peytcheva (2020) demonstrated that survey language influences comprehension, retrieval of relevant experiences, and response formation, while Aizpurua (2020) emphasized the importance of pretesting for identifying linguistic and cultural challenges that affect measurement quality. These considerations support the DCGM's integration of the Theory of Planned Behavior (TPB), which highlights the influence of beliefs and evaluations on attitudes and behavioral intentions (Ajzen, 1991). By preserving conceptual meaning in culturally appropriate Visayan, the refinement process sought to enable stakeholders to accurately convey their perceptions and experiences with local health initiatives. The observed stronger item loadings and improved measurement outcomes align with this approach, although the specific effect of linguistic refinement on model fit was not isolated.

The favorable SEM-based validation results reflect the combined influence of theoretical specification, systematic instrument development, and empirical refinement. Participatory linguistic and cultural adaptation was essential in improving respondents' comprehension and interpretation of the intended constructs. The observed increases in factor loadings following refinement provide empirical support for recognizing language as a consequential factor in measurement quality. Consequently, the DCGM highlights the importance of linguistic and cultural appropriateness in the development and validation of survey instruments, particularly for application in culturally and linguistically diverse communities.

Beyond its methodological contribution, the DCGM's significance extends to its substantive application in program evaluation and governance. The results support the DCGM structure illustrated in Figure 1. The Context, Input, Process, and Product dimensions represent interconnected yet distinct domains of program evaluation. Together, these dimensions reflect a program's operational flow: identifying community needs and priorities, mobilizing resources and institutional support, implementing activities and services, and assessing perceived

outcomes and sustainability. As each dimension provides unique information, the DCGM enables program planners and implementers to identify perceived strengths and areas requiring further examination.

Stakeholder perception is essential to the evaluative process, as program effectiveness is both experienced and documented administratively. Stakeholders interact with an initiative based on its relevance to community needs, accessibility of resources and services, implementation, and the outcomes they experience or observe. TPB offers a behavioral rationale for considering these perceptions consequential, since beliefs and evaluations influence attitudes, intentions, and subsequent engagement (Ajzen, 1991). In line with Davis (1989), subjective appraisals constitute meaningful evidence, even when not directly aligned with objective conditions. McCormick (1981) also identified acceptability and responsiveness to perceived needs as legitimate indicators of effectiveness. Within the DCGM, these perceptions serve as a feedback mechanism between communities and local government. The relationship labeled "influences perception" in Figure 1 indicates that program experiences shape stakeholders' subsequent assessments and engagement, allowing their feedback to inform ongoing evaluation.

This feedback informs governance through the relationship labeled "guides decision-making" in Figure 1. Distinct dimension scores enable planners and implementers to determine whether concerns pertain to program relevance, resources and institutional support, implementation and service delivery, or perceived outcomes and sustainability. Instead of relying solely on general judgments of success or failure, decision-makers can identify specific areas requiring attention and implement targeted adjustments to planning, resource allocation, coordination, or service delivery. Findings related to Product and stakeholder perceptions may prompt renewed assessment of Context, reconsideration of Input, or corrective measures within Process. This ongoing cycle of evidence, feedback, decision-making, action, and reassessment demonstrates the DCGM's dynamic nature. The approach is consistent with effectiveness-oriented management, where actual results inform subsequent decisions and actions (Drucker, 2006), and with McCormick's (1981) emphasis on evaluation as a foundation for improving practice. As a result, the question *Are we doing it right?* becomes a continuous aspect of program monitoring, accountability, and improvement.

Perception-based evidence serves to complement, not replace, administrative records, financial data, accomplishment reports, service-delivery metrics, or objective program outcomes. Recent research on instrument development has similarly distinguished perceived competence from demonstrated practice and recommended interpreting perception scores in conjunction with other evidence (Ertay & Gilanlioglu, 2024). The DCGM's unique contribution lies in its capacity to reveal whether stakeholders perceive an initiative as relevant, adequately supported, effectively implemented, and beneficial. When combined with administrative and performance records, these assessments provide local government leaders, planners, and implementers with a more comprehensive basis for understanding program challenges, prioritizing interventions, and making evidence-based decisions.

Although this validation focused on local health initiatives, the DCGM's functional logic may inform evaluation in other program contexts where stakeholder feedback is significant. Researchers and institutions may adapt its indicators to align with the objectives, target populations, languages, and operational conditions of different initiatives, while maintaining the integrated functions of CIPP, perception, feedback, and effectiveness-oriented decision-making. Any adaptation would require renewed participatory content review and empirical validation within the intended context. Future research may also utilize the instrument at various stages of the program cycle to examine how dimension scores relate to participation, service utilization, administrative performance, and objective outcomes. Such studies can further clarify the role of stakeholder feedback in promoting adaptive, accountable, and responsive governance.

Conclusion

This study concludes that the DCGM Evaluation Tool provides a reliable and construct-valid means of assessing the perceived effectiveness of local health initiatives across four related yet distinct dimensions: Context, Input, Process, and Product. These dimensions enable perceived program effectiveness to be examined through differentiated evidence concerning responsiveness to community needs, adequacy of resources and institutional support, implementation and service delivery, and perceived outcomes and sustainability.

The principal contribution of the DCGM lies in translating the CIPP framework, the behavioral significance of stakeholder perceptions, and effectiveness-oriented decision-making into a culturally and linguistically adapted evaluation instrument**.** Its development further demonstrates the importance of participatory linguistic and cultural refinement as a consequential consideration in establishing measurement quality, particularly in community-based evaluations involving linguistically diverse populations. By systematically returning stakeholder experiences to the evaluation cycle, the instrument embodies the feedback dynamics required for responsive local governance. It provides local governments, program planners, implementers, and community stakeholders with a common and structured basis for examining the question, *Are we doing it right?* and using the resulting evidence to inform continuing program review, decision-making, adjustment, and improvement.

A principal policy implication of the DCGM is the proposed establishment of a Municipal Evaluation Policy Framework (MEPF) with objectives that include institutionalizing systematic evaluation of municipal programs, projects, and activities (PPAs), enhancing transparency, accountability, and evidence-based decision-making in public resource utilization, and establishing continuous feedback mechanisms to improve municipal interventions. The development of the MEPF should involve a collaborative process with key decision-makers, program and project designers, implementers, barangay-level administrators, and the Local Finance Committee. This inclusive approach is anticipated to promote institutional ownership, facilitate cooperation across governance levels, and ensure that evaluation procedures align with relevant financial and administrative requirements.

The proposed MEPF may adopt a structured evaluation cycle that includes, at a minimum, an inventory of existing PPAs, prioritization of evaluations, evaluation planning, evidence collection and analysis, management and policy responses, and ongoing monitoring and reassessment. For new PPAs requiring financial resources, proposals submitted to the Local Chief Executive should incorporate an evaluation component specifying objectives, indicators, activities, responsibilities, schedules, and expected deliverables. The DCGM may be utilized to gather stakeholder-perception evidence in conjunction with administrative and performance records to inform planning, budgeting, and implementation decisions. As a dynamic policy framework, the MEPF remains adaptable to additional objectives, components, and procedural stages identified through participatory development and alignment with municipal governance requirements.

The present validation supports the use of the four dimension scores within the context studied. These scores provide perception-based evidence that can complement administrative records, service-delivery measures, and objective program outcomes. Although the DCGM may inform the evaluation of other local initiatives, applications beyond the present health context require appropriate contextual adaptation and renewed validation.

Contributions of Author

The sole author conceptualized and designed the study, developed the research instrument, coordinated data collection, conducted the statistical analyses, interpreted the findings, and prepared, reviewed, and approved the final manuscript.

Funding

The study received institutional support from the Municipality of Tudela, Cebu, Philippines, with existing office resources. Financial support was allocated specifically for the journal publication fee, subject to reimbursement upon submission of the official receipt. No other funds from the approved research budget were disbursed.

Conflict of Interests

The author served as Municipal Administrator of Tudela, Cebu, during the conduct of the study and was designated as convenor of the research activity team under Executive Order No. 19-B, Series of 2024. This institutional relationship is disclosed for transparency. The author declares no financial conflict of interest.

Acknowledgment

The author gratefully acknowledges the Local Government Unit of Tudela, Cebu, and the participating community for their support of this study. Special appreciation is extended to Hon. Greman B. Solante, Municipal Mayor, whose endorsement and encouragement made the research possible. The author also thanks the department heads and personnel of the Municipal Social Welfare and Development Office, led by Jemerson A. Solante, MPA, and the Rural Health Unit, under the leadership of Maria Eva B. Daclan, MD, Municipal Health Officer. Their coordination and field support facilitated data collection and engagement with the study participants.

The author likewise acknowledges Pablito dela Rama, PhD, for his guidance on the instrument-validation design and quantitative analysis, and Alice Mae M. Arbon, PhD, for her linguistic expertise during the cultural and linguistic adaptation of the instrument from English into the local variety of Visayan used in Tudela. Their respective contributions strengthened the methodological and linguistic foundations of the study.

Declaration on the Use of AI-assisted Technologies

During manuscript preparation and finalization, the author used Google Search's AI Overview, ChatGPT, and Grammarly as assistive tools for language refinement, organization, and clarity of presentation. These tools supported the articulation and structuring of the author's ideas and did not replace the author's scholarly judgment. The author reviewed and edited all AI-

assisted material, treated it only as guidance, and assumes full responsibility for the accuracy, interpretation, originality, and final content of the manuscript.

References

Acenas, A. (2023). Evaluation of undergraduate research course using Stufflebeam's context, input, process, product model. Journal of Interdisciplinary Perspectives, 1(1), 60–69. https://doi.org/10.69569/jip.2023.0003

Aizpurua, E. (2020). Pretesting methods in cross-cultural research. In M. Sha & T. Gabel (Eds.), The Essential Role of Language in Survey Research (pp. 129–150). RTI International. https://doi.org/10.3768/rtipress.bk.0023.2004

Ajzen, I. (1991). The theory of planned behavior. Organizational Behavior and Human Decision Processes, 50(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Ajzen, I., & Fishbein, M. (2014). The influence of attitudes on behavior. In D. Albarracin, B. T. Johnson, & M. P. Zanna (Eds.), The Handbook of Attitudes (pp. 173–221). Psychology Press. https://doi.org/10.4324/9781410612823 (Original work published 2005, Lawrence Erlbaum Associates Publishers)

Altarriba, J. (1992). The representation of translation equivalents in bilingual memory. Advances in Psychology, 83(C), 157–174. https://doi.org/10.1016/S0166-4115(08)61493-4

Amoto, J., Sodusta, D.J., Fernandez, Jr., P., Firmase, J., & Nanta, E.M. (2024). Evaluating the University of the Philippines Visayas Action Research Project based on the context, input, process, and product model. Davao Research Journal, 15(4), 120–130. https://doi.org/10.59120/drj.v15i4.280

Balgos, B. (2023). NEPF summative evaluation of the National Evaluation Policy Framework—Final report (pp. 1–72) [Evaluation report]. Department of Economy, Planning, and Development. https://nep.depdev.gov.ph/evaluations/105

Betsch, C., Böhm, R., & Chapman, G. (2015). Using behavioral insights to increase vaccination policy effectiveness. Policy Insights from the Behavioral and Brain Sciences, 2(1), 61–73. https://doi.org/10.1177/2372732215600716

Catubig, R.M.C. (2026). School-based nutrition programs in Philippine basic education: A context input process and product (CIPP)-based evaluation model. International Journal of the Humanities and Social Sciences, 5(2), 5860–5887. https://doi.org/10.69651/2022PIJHSS

Çelebi, I., Gökkaya, E., Çopur, A., & Balci, E. (2025). Quiet quitting reason and behavior scale design and development process. Work, 80(1), 164–172. https://doi.org/10.3233/WOR-230699

Corpus, V. (2022). The Philippines' National Evaluation Policy Framework, evaluation practice, challenges and prospects [Policy framework presentation document]. National Economic and Development Authority. https://tinyurl.com/yfctxrp6

Davis, F. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13(3), 319–338. https://doi.org/10.2307/249008

dela Luna, K.L., Sy, A.D., Sabandal, R.L.C., Jacinto, C.B., Malimban, R., Sales, B.E., Marollano, R., & Marin, A. (2026). A qualitative program evaluation study on the perceived impact of health and nutrition programs among beneficiaries of a civil society organization in the Philippines. Acta Medica Philippina. https://doi.org/10.47895/amp.vi0.12745

Ding, J., Wang, A., & Zhang, Q. (2023). Mining the vaccination willingness of China using social media data. International Journal of Medical Informatics, 170, 104941. https://doi.org/10.1016/j.ijmedinf.2022.104941

Drucker, P.F. (2006). The effective executive: The definitive guide to getting the right things done. Harper Business.

Ertay, Z., & Gilanlioglu, I. (2024). The development and validation of an intercultural communicative competence scale for students. SAGE Open, 14(2), 1–11. https://doi.org/10.1177/21582440241252495

Fornell, C., & Larcker, D. (1981). Evaluating structural equation models with unobservable variables and measurement error. Journal of Marketing Research, 18(1), 39–50. https://doi.org/10.1177/002224378101800104

Hair, J.F. Jr., Hult, G.T.M., Ringle, C.M., & Sarstedt, M. (2022). A primer on partial least squares structural equation modeling (PLS-SEM) (Third Edition). SAGE Publications, Inc.

Henseler, J., Ringle, C., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. Journal of the Academy of Marketing Science, 43(1), 115–135. https://doi.org/10.1007/s11747-014-0403-8

Khoury, B., & Vergara, R. (2026). Compassion questionnaires revised: Scales development and validation. Assessment, 33(4), 653–687. https://doi.org/10.1177/10731911251337185

Kline, R.B. (2016). Principles and practice of structural equation modeling (4th Ed.). The Guilford Press.

Luib, S., Loquias, M., Capule, F., Garcia, K.E., & Sendito, M.J. (2022). Evaluation of the patient medication counseling services in the Philippine General Hospital using the CIPP model. Acta Medica Philippina, 56(6), 143–150. https://doi.org/10.47895/amp.v56i6.4522

McCormick, J.S. (1981). Effectiveness and efficiency. Journal of the Royal College of General Practitioners, 31(226), 299–302. https://pmc.ncbi.nlm.nih.gov/articles/PMC1971043/

Metehan, C.T., Sözer, E.G., Civelek, M.E., Ertemel, A.V., & Mert, I.S. (2025). Buyer platform experience (PLX) in online marketplaces: Scale development, validation, and the salience of variety, competitive pricing, service excellence, and promotion. SAGE Open, 15(4), 1–20. https://doi.org/10.1177/21582440251394446

Municipality of Tudela. (2024). "Are we doing it right?": Decoding local health initiatives through the lens of program effectiveness evaluation [Approved research activity proposal]. Office of the Municipal Administrator, Municipality of Tudela.

National Economic and Development Authority [NEDA]. (2020). Guidelines on evaluation in the national government (pp. 1–94) [Guidelines document]. NEDA.

Nunnally, J.C. (1978). Psychometric theory (2nd Edition). McGraw-Hill, Inc.

Office of the Mayor of Tudela. (2024). Executive Order No. 19-B series of 2024: An order creating and mandating an activity team to conduct the scientific study "Are we doing it right?": Decoding local health initiatives through the lens of program effectiveness (Executive Order No. 19-B Series of 2024). Municipality of Tudela.

Peytcheva, E. (2020). The effect of language of survey administration on the response formation process. In M. Sha & T. Gabel (Eds.), The Essential Role of Language in Survey Research (pp. 3–21). RTI International. https://doi.org/10.3768/rtipress.bk.0023.2004

Phang, G., & Ilham, Z. (2023). Theory of planned behavior to understand pro-environmental behavior among Universiti Malaya students. AIMS Environmental Science, 10(5), 691–708. https://doi.org/10.3934/environsci.2023038

Prasetyo, Y.T., Castillo, A.M., Salonga, L.J., Sia, J.A., & Seneta, J.A. (2020). Factors affecting perceived effectiveness of COVID-19 prevention measures among Filipinos during enhanced community quarantine in Luzon, Philippines: Integrating protection motivation theory and extended theory of planned behavior. International Journal of Infectious Diseases, 99, 312–323. https://doi.org/10.1016/j.ijid.2020.07.074

Repke, L., & Felderer, B. (2025). The relationship between measurement error, representation bias, language, and country: A comparative analysis using the European Social Survey (Rounds 5 to 7). Survey Methods: Insights from the Field, Special Issue 'Advancing Comparative Research: Exploring Errors and Quality Indicators in Social Research.' https://doi.org/10.13094/SMIF-2025-00013

Rivera, A.L.S., & Altarriba, J. (2002). The role of language in therapy with the Spanish-English bilingual client. Professional Psychology: Research and Practice, 33(1), 30–38. https://doi.org/10.1037/0735-7028.33.1.30

Rodrigues, F., Ziade, N., Jatuworapruk, K., Caballero-Uribe, C.V., Khursheed, T., & Gupta, L. (2023). The impact of social media on vaccination: A narrative review. Journal of Korean Medical Science, 38(40), e326. https://doi.org/10.3346/jkms.2023.38.e326

Rossi, P.H., Lipsey, M.W., & Freeman, H.E. (2004). Evaluation: A systematic approach (7th Ed.). SAGE Publications, Inc.

Sankaran, S., & Saad, N. (2022). Evaluating the Bachelor of Education program based on the context, input, process, and product model. Frontiers in Education, 7, 924374. https://doi.org/10.3389/feduc.2022.924374

Seddig, D., Maskileyson, D., Davidov, E., Ajzen, I., & Schmidt, P. (2022). Correlates of COVID-19 vaccination intentions: Attitudes, institutional trust, fear, conspiracy beliefs, and vaccine skepticism. Social Science & Medicine, 302, 114981. https://doi.org/10.1016/j.socscimed.2022.114981

Social Weather Stations [SWS]. (2025). Fourth Quarter 2024 Social Weather Survey: Self-rated poverty (SRP) at 63% in December 2024, up from 59% in September; borderline at 11%, not poor at 26% [Social weather report]. Social Weather Station. https://sws.org.ph/social-weather-report-fourth-quarter-2024-social-weather-survey-self-rated-poverty

Stufflebeam, D. (2003). The CIPP model for evaluation. In T. Kellaghan, D. L. Stufflebeam, & L. A. Wingate (Eds.), International Handbook of Educational Evaluation Part 1 (Vol. 9, pp. 31–62). Kluwer Academic Publishers. https://doi.org/10.1007/978-94-010-0309-4_4

Sundari, S. (2019). Implementation CIPP model for exploring evaluation of analysis performance policy to Ministry Defense Indonesia (MDI). International Journal of Scientific & Technology Research, 8(10), 1417–1423. https://tinyurl.com/3xe866yu

Tabachnick, B.G., & Fidell, L.S. (2014). Using multivariate statistics (Sixth Edition, Pearson New International Edition). Pearson.

United Nations General Assembly. (2015). Transforming our world: The 2030 agenda for sustainable development (General Assembly Resolution A/RES/70/1). United Nations. https://digitallibrary.un.org/record/3923923/

U.S. Government Accountability Office [GAO]. (2021). Program evaluation: Key terms and concepts. U.S. Government Printing Office. https://www.gao.gov/assets/gao-21-404sp.pdf

Wahjusaputri, S., & Irawan, B. (2019). Bureaucratic reform policy strategy using the CIPP model for the improvement of public services in Kemenko-Polhukam. Jurnal Ekonomi Dan Bisnis, 3(1), 1–16. https://journal.uhamka.ac.id/index.php/agregat/article/view/3013

Wandani, F.T., Widodo, A., Rubiyanti, Nurafni, Rubiyanti, N., & Silvianita, A. (2024). Analysis of the influence of environmental concern, theory of planned behavior, and theory of reasoned action on purchase intention: A conceptual paper. International Journal of Scientific Multidisciplinary Research, 2(5), 499–508. https://doi.org/10.55927/ijsmr.v2i5.9138

Zhang, G., Zeller, N., Griffith, R., Metcalf, D., Williams, J., Shea, C., & Misulis, K. (2011). Using the context, input, process, and product evaluation model (CIPP) as a comprehensive framework to guide the planning, implementation, and assessment of service-learning programs. Journal of Higher Education Outreach and Engagement, 15(4), 57–83. https://openjournals.libs.uga.edu/jheoe/article/view/901

Zhang, Y., & Aryadoust, V. (2022). A systematic review of the validity of questionnaires in second language research. Education Sciences, 12(723), 1–32. https://doi.org/10.3390/educsci12100723