

## Research Article

### BENEFICIARY SATISFACTION WITH A COAL MINING COMPANY'S STINGLESS BEE PROGRAM IN NORTH BARITO, INDONESIA

<sup>1,\*</sup> Efredi Yudianto, <sup>2</sup> Arianto Sasono, <sup>2</sup> Rapinis Mutiara, <sup>1</sup> Muhammad Rum, <sup>2</sup> Abdul Syukur, <sup>2</sup> Bagus M.M, <sup>2</sup> Muhammad Rizky N, <sup>2</sup> Rayhan Rajoalam Putra P, <sup>2</sup> Nurlaila Khoirunnisa, <sup>3</sup> Dewi Hajar

<sup>1</sup>PT Tuah Turangga Agung, Jakarta, Indonesia.

<sup>2</sup>PT Suprabari Mapanindo Mineral, North Barito Regency, Central Kalimantan, Indonesia.

<sup>3</sup>Department of Environmental Engineering, Diponegoro University, Semarang, Central Java, Indonesia.

Received 10<sup>th</sup> August 2026; Accepted 11<sup>th</sup> September 2026; Published online 25<sup>th</sup> October 2026

#### ABSTRACT

**Aims:** To evaluate beneficiary satisfaction with GEMA MADU (Gerakan Masyarakat Budidaya Madu Kelulut), a stingless bee (kelulut) honey empowerment program of a coal mining company, using the Community Satisfaction Index (IKM), and to translate item-level results into management priorities. **Study Design:** Cross-sectional survey with secondary quantitative analysis. **Place and Duration of Study:** Lemo I Village, Teweh Tengah District, North Barito Regency, Central Kalimantan, Indonesia, between July and August 2026. **Methodology:** Sixteen program beneficiaries rated 17 items on a four-point scale covering the overall program, program management, and service delivery. Rating counts were reconstructed from published percentages and validated against the reported management index. Each item was tested against the satisfaction threshold of 3.00 with Holm-adjusted one-sample t tests and exact sign tests, the indices were graded under three interpretation schemes, and items were assigned to five action tiers. **Results:** The reported indices (overall 74.116, management 74.609, service delivery 72.545) were "satisfied" on the report scale, grade B under the 2004 guideline, and grade C under the 2017 guideline. Facilitator competence scored highest (mean 3.56, 95% CI 3.29–3.84; adjusted P = .009) and was the only robust strength. Method appropriateness and fairness of targeting scored lowest (2.50; adjusted P = .02) under a stated reconstruction assumption. The benefit items, including income (2.75), did not differ significantly from the threshold. Raising the five weakest items to the threshold would require 33 one-step improvements among 80 ratings. **Conclusion:** Beneficiaries value the facilitator, whereas the effectiveness of the program design has not yet been demonstrated. The company should redesign and publish the targeting rules, record production and income, plan a staged exit, and fix the grading scheme and report item-level confidence intervals in future surveys.

**Keywords:** Community satisfaction index; corporate social responsibility; stingless bee; coal mining; program evaluation.

## INTRODUCTION

### Corporate Community Programs in the Mining Sector

Mining companies in Indonesia are expected to extract resources efficiently and, at the same time, to improve the welfare of the communities that live beside their operations. This expectation is also a legal obligation. Law No. 40 of 2007 makes corporate social responsibility (CSR) mandatory for companies whose business relates to natural resources [1,2], and sectoral regulation requires mineral and coal mining companies to plan, fund, and report community development and empowerment programs [3,4]. Beyond legal compliance, communities that perceive themselves as excluded from the benefits of extraction create socio-political risks that raise operating costs and weaken the social license to operate [5,6]. Community acceptance of mining depends on perceived procedural fairness, on the quality of contact between company and community, and on trust [7].

Community development programs nevertheless frequently fall short of their objectives. They tend to decline when the company withdraws, fail to address actual community needs, or lack systematic monitoring [2]. Geothermal CSR programs in West Java scored well on effectiveness and relevance but poorly on sustainability and empowerment [8], and stakeholders around a large nickel mine in South Sulawesi perceived that the company could demonstrate

greater concern despite substantial social investment [9]. Stakeholder theory explains why such perceptions matter, since an organization is accountable to every group affected by its activities [10], and stakeholder analysis enables mining companies to incorporate these expectations into program planning [11]. Beneficiary evaluation is therefore a central element of program governance rather than an administrative formality.

### Stingless Beekeeping as an Empowerment Strategy

Stingless beekeeping, or meliponiculture, is widely adopted in corporate empowerment programs. Stingless bees occur throughout tropical Indonesia [12] and produce honey, pot-pollen, and propolis in simple box hives that require little capital [13]. Yields, however, are modest. Colonies of *Tetragonula laeviceps* in Yogyakarta produced 60 to 263 ml of honey over four months [14], and 56 colonies of *T. clypearis* in West Lombok generated approximately USD 500 in eight months, or about USD 9 per colony [15]. Companies have combined hive distribution with a training center and retail outlet [16] or with blended face-to-face and messaging-group mentoring [17]. The degree to which beneficiaries are satisfied with such programs, and with which components, is therefore a practical management question with direct implications for program design.

### The Community Satisfaction Index and Its Limitations

Satisfaction research draws on expectation–disconfirmation theory [18] and its multidimensional extensions such as SERVQUAL [19]. Both frameworks hold that ratings reflect what people expected as

\*Corresponding Author: Efredi Yudianto,  
1PT Tuah Turangga Agung, Jakarta, Indonesia.

well as what they received [20], so that a group with modest expectations may rate a modest program highly. Indonesia standardized the Community Satisfaction Index (*Indeks Kepuasan Masyarakat*, IKM) for public-service units on the basis of Likert-type items [21], equal weights, and a conversion to a 25–100 scale [22,23]. Most published IKM studies examine public services. A public service mall in Banjarbaru scored 78.8 from 150 respondents [24], a licensing office in Tanjungpinang scored 66.49, with speed and schedule certainty weakest [25], an e-KTP service in Madiun scored approximately 80.7 [26], and a civil registry in Central Lombok scored 83.46 from 400 respondents [27]. In public services, the weakest elements therefore relate mainly to process attributes such as speed, schedules, and facilities.

The IKM has also been adapted to CSR programs, including red ginger cultivation [28], village development programs [29], a mining company's community development scheme [30], and a national park empowerment program [31]. These adaptations share three methodological limitations. First, they report indices to two or three decimals from small samples without quantifying uncertainty or testing hypotheses. Second, they apply grading schemes inconsistently, so that the same value may be classified as "satisfied", "good", or "fair". Third, they aggregate away the item-level differences that indicate priorities for improvement. The company report on which the present study is based follows the same pattern and provides index values and a narrative only.

## Objectives and Hypotheses

This study aims to evaluate the satisfaction of beneficiaries of GEMA MADU (*Gerakan Masyarakat Budidaya Madu Kelulut*, the Lemo I Village Stingless Bee Honey Cultivation Movement), a program run by PT Suprabari Mapanindo Mineral (SMM), a coal mining company in North Barito Regency, Central Kalimantan. The specific objectives are to reproduce and validate the reported indices, to quantify their uncertainty, to test hypotheses against the satisfaction threshold and national benchmarks, and to translate item-level results into action tiers for management. Four hypotheses were specified. H1 states that facilitator competence, facilitator responsibility, and company responsiveness exceed the satisfaction threshold (mean > 3.00). H2 states that method appropriateness, fairness of targeting, and participation in monitoring and evaluation fall below the threshold (mean < 3.00). H3 states that items describing the delivery team score higher than items describing program outcomes. H4 states that the management index falls short of the 2017 grade B boundary (76.61). H1 and H2 are primary, and H3 and H4 are secondary.

## METHODOLOGY

### Study Design and Setting

This study is a secondary quantitative analysis of a cross-sectional satisfaction survey conducted in Lemo I Village, Teweh Tengah District, North Barito Regency, adjacent to the operations of SMM. The Semarang branch of PT SUCOFINDO designed and administered the survey as part of the company's routine monitoring and evaluation (PT SUCOFINDO, unpublished survey report, 2026). A focus group discussion and structured questionnaire interviews took place in July and August 2026. The analysis used the aggregated results of the survey report, namely rating percentages, indices, and narrative, and no names or respondent-level records, in line with data-minimization principles [32]. The survey report belongs to the company and is available from the corresponding author on reasonable request, subject to the company's approval. Reporting follows the STROBE guidelines for cross-sectional studies where

applicable [33]. GEMA MADU is a company-funded stingless bee honey program facilitated by a Community Development Officer (CDO), and its beneficiaries take part in planning, implementation, and monitoring. The report provides no information on hive numbers, start date, budget, production, or sales.

### Participants and Sample Size

The survey interviewed 16 GEMA MADU beneficiaries, all residents of Lemo I, who were selected to represent the groups receiving company programs. The report states no eligibility criteria and no participation rate. Its respondent list repeats two first names with different address formats; these may be namesakes, and  $N = 16$  is used as reported. Because the sample is purposive, it describes this beneficiary group and does not support statistical generalization to the village. With  $N = 16$ , an unadjusted two-sided  $t$  test at  $\alpha = .05$  reaches 80% power only for standardized effects  $|d_z| \geq 0.75$ , and the Holm correction across 17 tests raises this threshold to approximately 1.1. A non-significant item therefore indicates that the data are inconclusive for that item rather than that beneficiaries are satisfied with it.

### Instrument and Index Computation

The closed questions used a four-point Likert-type scale (1 = not satisfied, 2 = less satisfied, 3 = satisfied, 4 = very satisfied), with labels adapted to each question. They covered three dimensions: the overall program (relevance, four benefits, and continuity); program management (participation in planning, implementation, and monitoring and evaluation; stakeholder synergy; and four attributes of the CDO); and service delivery (method appropriateness, reliability, fairness of targeting, and responsiveness). Open-ended questions asked respondents to explain their ratings. The report provides no psychometric evidence for the instrument. For question  $q$  answered by  $N$  respondents, the mean score is

$$M_q = (1/N) \sum x_{iq} \quad (1)$$

where  $x_{iq}$  is the rating of respondent  $i$ . With equal weights of  $1/K_d$  for the  $K_d$  questions in dimension  $d$ , the dimension index and its converted value are

$$CSI_d = \sum_{q \in d} (1/K_d) M_q, \quad IKM_d = 25 \times CSI_d \quad (2)$$

so that the lowest possible score of 1 maps to 25 and the highest possible score of 4 maps to 100 [22,23].

### Reconstruction and Validation of Rating Counts

The report charts the rating shares of 17 items: six for the overall program, eight for management, and three for delivery. With  $N = 16$ , each percentage corresponds to a whole number of respondents (6.25% per person), so the shares were converted exactly into counts of ratings 2, 3, and 4. No respondent selected rating 1. The reconstruction was validated on the only fully charted dimension: the eight management items average 2.984 (IKM 74.609), which matches the reported value exactly. Two items required an assumption and are flagged in Tables 3 and 5. For method appropriateness, the report states that 50% of respondents judged the method appropriate, and the remaining half was assigned rating 2. For fairness, it states that 50% found targeting less appropriate, and the remaining half was assigned rating 3; the extreme alternative, in which all of them rated 4, was examined as a sensitivity case.

## Statistical Analysis

Each item was tested against the satisfaction threshold of 3.00 with a one-sample  $t$  test ( $df = 15$ ). The 95% confidence interval and the standardized effect were computed as

$$M_q \pm t_{0.975,15} \cdot s_q / \sqrt{N}, \quad d_z = (M_q - 3) / s_q \quad (3)$$

where  $t_{0.975,15} = 2.131$  and  $s_q$  is the sample standard deviation. Values of  $|d_z|$  near 0.2, 0.5, and 0.8 are conventionally interpreted as small, medium, and large effects [34]. Because ratings take only three values, an exact two-sided sign test was added; it is equivalent to the Wilcoxon signed-rank test here because all non-zero deviations have the same magnitude. Means of ordinal items support  $t$ -based inference when used with care [35,36], and estimates with intervals are emphasized over isolated  $P$  values [37]. Seventeen tests create a multiplicity problem, so  $P$  values were adjusted with Holm's sequential procedure [38]. An item was classified as a robust strength or weakness when its Holm-adjusted  $P$  value was below .05, and as a nominal strength or weakness when only its unadjusted 95% interval excluded 3.00. Shares of respondents rating 2 are reported with Wilson score intervals, which have better coverage than Wald intervals in small samples [39,40].

For H3, the five delivery-team items were compared with the six outcome items using an exact Mann–Whitney test on item means. For H4 and the external benchmarks, the mean of item means in each group was tested by one-sample  $t$  tests with  $k - 1$  degrees of freedom against four values: 75.00 (the report's "very satisfied" cut-off), 76.61 (2017 grade B), 65.00 (the 2017 grade C floor), and 84.75 (a red ginger CSR program [28]). Treating items as units describes the consistency of items rather than the sampling of respondents, so these tests are exploratory; the 12 benchmark tests were adjusted jointly with Holm's method. All computations were performed in Python with SciPy.

## Interpretation Schemes and Action Tiers

Three grading schemes were applied to the converted indices (Table 1). The report divides the range into equal bands, but its lowest band (0–25) cannot occur because the minimum converted score is 25. The 2004 guideline sets grade boundaries at 43.75, 62.50, and 81.25 [22], and the 2017 guideline raises them to 65.00, 76.60, and 88.30 [23].

**Table 1. Interpretation schemes and the placement of dimensions and items**

| Scheme               | Bands (converted IKM) <sup>a</sup>                             | Dimensions (n = 3) | Items (n = 17)                   |
|----------------------|----------------------------------------------------------------|--------------------|----------------------------------|
| Report scale         | Less satisfied 25–<50; satisfied 50–<75; very satisfied 75–100 | Satisfied: 3       | Satisfied: 11; very satisfied: 6 |
| KepMenPAN 25/2004    | D 25–43.75; C 43.76–62.50; B 62.51–81.25; A 81.26–100          | B: 3               | A: 3; B: 12; C: 2                |
| PermenPAN-RB 14/2017 | D 25–64.99; C 65.00–76.60; B 76.61–88.30; A 88.31–100          | C: 3               | A: 1; B: 4; C: 10; D: 2          |

<sup>a</sup>Boundary values follow the stated cut-offs: 62.50 is grade C in 2004, and 75.00 is "very satisfied" on the report scale. The lowest band of the report scale (0–25) cannot occur because the converted minimum is 25.

To translate the tests into decisions, each item was assigned to one of five action tiers using only the item statistics (Table 3). Tier A (sustain) holds items whose lower 95% bound exceeds 3.00, and tier B (correct) holds items whose upper bound lies below 3.00. For the remaining items, whose interval contains 3.00, the share of respondents rating 2 determines the tier: tier C (verify) when at least one third of respondents (6 of 16) rated 2, tier D (monitor) when one to five did so, and tier E (maintain) when none did. The one-third cut-off is a convention rather than a statistical result, and its sensitivity is reported. For items below 3.00, the number of one-step rating improvements required to raise the mean to the threshold was calculated as

$$U_q = \lceil N(3 - M_q) \rceil \quad (4)$$

Which expresses the size of the change in beneficiaries' experience needed to close the gap.

## RESULTS AND DISCUSSION

### Respondent Characteristics

Most respondents were young or middle-aged: five (31%) were 25 years or younger, nine (56%) were 26–45 years, and two (13%) were 46–65 years. Eight (50%) had completed junior high school, four (25%) senior high school, three (19%) primary school, and one (6%) had no formal schooling. Two respondents (12.5%) were self-employed, and 14 (87.5%) worked in farming, labor, household management, fishing, or other occupations. Nine (56.25%) earned less than IDR 1,000,000 per month, and seven (43.75%) earned IDR 1,000,000–3,000,000. The report provides no refusal data.

### Dimension Indices and Interpretation Schemes

The overall program scored 2.965 (IKM 74.116), program management 2.984 (74.609), and service delivery 2.902 (72.545) (Table 2, Fig. 1). All three dimensions are "satisfied" on the report scale, grade B ("Good") under the 2004 guideline, and grade C ("Fair") under the 2017 guideline. The difference between the highest and lowest dimension is only 2.06 points, which is too small to distinguish the dimensions with 16 respondents. The reported total (2.965; 74.116) is identical to the overall-program index, which indicates that the total was carried over from that dimension rather than computed across all three. Two alternative composites were therefore calculated: the unweighted mean of the dimensions (2.950; 73.76) and the mean of the 17 charted items (2.912; 72.79). The three composites lie within 1.4 points of one another and receive the same grade under every scheme.

Table 2. Dimension indices, composite variants, and distance to thresholds

| Dimension or composite        | Index (1–4) | IKM    | Report scale | 2004 <sup>a</sup> | 2017 <sup>a</sup> | Points below 75.00 | Points below 76.61 |
|-------------------------------|-------------|--------|--------------|-------------------|-------------------|--------------------|--------------------|
| Overall program <sup>b</sup>  | 2.965       | 74.116 | Satisfied    | B                 | C                 | 0.88               | 2.49               |
| Program management            | 2.984       | 74.609 | Satisfied    | B                 | C                 | 0.39               | 2.00               |
| Service delivery <sup>b</sup> | 2.902       | 72.545 | Satisfied    | B                 | C                 | 2.46               | 4.07               |
| Reported total <sup>c</sup>   | 2.965       | 74.116 | Satisfied    | B                 | C                 | 0.88               | 2.49               |
| Unweighted mean of dimensions | 2.950       | 73.76  | Satisfied    | B                 | C                 | 1.24               | 2.85               |
| Mean of 17 charted items      | 2.912       | 72.79  | Satisfied    | B                 | C                 | 2.21               | 3.82               |

<sup>a</sup>Grade B under the 2004 guideline is labeled “Good”, and grade C under the 2017 guideline is labeled “Fair”.

<sup>b</sup>Reported values that cannot be reproduced from the charted items (Section 3.5).

<sup>c</sup>Identical to the overall-program index in the survey report.

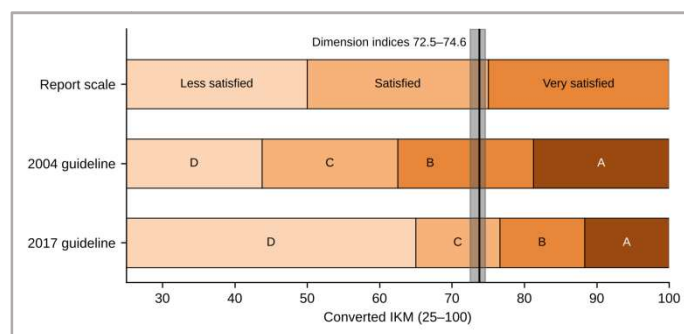


Fig. 1. Converted IKM of the three dimensions against three interpretation schemes

Shaded band: range of the three dimension indices (72.5–74.6); vertical line: their unweighted mean (73.76)

The distance to the next grade depends on the scheme. The management index lies 0.39 points below 75.00; its total rating score would need to rise from 382 to 384 of 512 points, which corresponds to two respondent-item ratings moving up one step. Reaching 2017 grade B (76.61) would require 11 such upgrades, or 8.6% of the 128 management ratings. The lenient classification (“very satisfied”) therefore depends on very few responses, whereas the stringent classification (2017 “Good”) does not.

### Item-Level Satisfaction

Three of the 17 items have 95% intervals entirely above 3.00 (Table 3): facilitator competence (mean 3.56, 95% CI 3.29–3.84; nine respondents rated it 4), company responsiveness (3.38; 3.11–3.64), and facilitator responsibility (3.31; 3.06–3.57). No respondent rated any of these items below satisfactory. Five items have intervals entirely below 3.00: method appropriateness and fairness (2.50; 2.22–2.78; 50% rated 2), participation in monitoring and evaluation and stakeholder synergy (2.63; 2.36–2.89; 37.5% rated 2), and facilitator closeness (2.69; 2.43–2.94; 31.3% rated 2). The Wilson interval for a 50% share extends from 28% to 72%, so the sample cannot distinguish between “half dissatisfied” and “a majority dissatisfied”. The remaining nine items, including relevance and all four benefit items, have intervals that contain 3.00. Among them, relevance (2.69), income (2.75), and knowledge (2.81) have the lowest means, with 38–44% of respondents rating them 2, whereas continuity (3.19) and planning participation (3.13) show a positive tendency.

Table 3. Item-level descriptive statistics and tests against the satisfaction threshold of 3.00 (N = 16)

| Item                                  | Ratings 2/3/4 <sup>a</sup> | M    | SD   | 95% CI    | d <sub>z</sub> <sup>b</sup> | Holm-adjusted P <sup>c</sup> |
|---------------------------------------|----------------------------|------|------|-----------|-----------------------------|------------------------------|
| <b>Overall program</b>                |                            |      |      |           |                             |                              |
| Relevance to needs                    | 7/7/2                      | 2.69 | 0.70 | 2.31–3.06 | –0.44                       | .77                          |
| Benefit, knowledge                    | 6/7/3                      | 2.81 | 0.75 | 2.41–3.21 | –0.25                       | >.99                         |
| Benefit, skills                       | 5/7/4                      | 2.94 | 0.77 | 2.53–3.35 | –0.08                       | >.99                         |
| Benefit, environment                  | 4/10/2                     | 2.88 | 0.62 | 2.55–3.20 | –0.20                       | >.99                         |
| Benefit, income                       | 7/6/3                      | 2.75 | 0.77 | 2.34–3.16 | –0.32                       | >.99                         |
| Continuity                            | 0/13/3                     | 3.19 | 0.40 | 2.97–3.40 | 0.47                        | .74                          |
| <b>Program management<sup>d</sup></b> |                            |      |      |           |                             |                              |
| Participation, planning               | 0/14/2                     | 3.13 | 0.34 | 2.94–3.31 | 0.37                        | >.99                         |
| Participation, implementation         | 4/8/4                      | 3.00 | 0.73 | 2.61–3.39 | 0.00                        | >.99                         |
| Participation, M&E                    | 6/10/0                     | 2.63 | 0.50 | 2.36–2.89 | –0.75                       | .13                          |
| Stakeholder synergy                   | 6/10/0                     | 2.63 | 0.50 | 2.36–2.89 | –0.75                       | .13                          |
| CDO closeness                         | 5/11/0                     | 2.69 | 0.48 | 2.43–2.94 | –0.65                       | .22                          |
| CDO adaptability                      | 1/15/0                     | 2.94 | 0.25 | 2.80–3.07 | –0.25                       | >.99                         |
| CDO responsibility                    | 0/11/5                     | 3.31 | 0.48 | 3.06–3.57 | 0.65                        | .22                          |
| CDO competence                        | 0/7/9                      | 3.56 | 0.51 | 3.29–3.84 | 1.10                        | .009                         |
| <b>Service delivery</b>               |                            |      |      |           |                             |                              |
| Method appropriateness <sup>e</sup>   | 8/8/0                      | 2.50 | 0.52 | 2.22–2.78 | –0.97                       | .02                          |
| Fairness <sup>e</sup>                 | 8/8/0                      | 2.50 | 0.52 | 2.22–2.78 | –0.97                       | .02                          |
| Responsiveness                        | 0/10/6                     | 3.38 | 0.50 | 3.11–3.64 | 0.75                        | .13                          |

<sup>a</sup>Number of respondents rating 2 (less satisfied), 3 (satisfied), and 4 (very satisfied); no respondent rated 1. Counts are reconstructed from reported percentages.

<sup>b</sup>d<sub>z</sub> = (M – 3)/SD.

<sup>c</sup>One-sample t test against 3.00 (df = 15), adjusted across 17 items.

<sup>d</sup>CDO = Community Development Officer; M&E = monitoring and evaluation.

<sup>e</sup>Assumption-dependent items (Section 2.4).

## Hypothesis and Benchmark Tests

H1 is supported for facilitator competence, which remains significant after Holm correction ( $t = 4.39$ , adjusted  $P = .009$ ,  $d_z = 1.10$ ). Responsiveness (adjusted  $P = .13$ ) and responsibility (adjusted  $P = .22$ ) show large effects ( $d_z = 0.75$  and  $0.65$ ) that are significant only before correction. H2 is supported for method appropriateness and fairness ( $t = -3.87$ , adjusted  $P = .02$ ,  $d_z = -0.97$ ), provided that the reconstruction assumption holds. Participation in monitoring and evaluation, stakeholder synergy (adjusted  $P = .13$  each), and closeness (adjusted  $P = .22$ ) are nominal weaknesses only. The exact sign test is more conservative, and no item remains significant after correction (smallest adjusted  $P = .07$ , facilitator competence). Nine items cannot be distinguished from 3.00, and their effects are small to medium ( $|d_z| \leq 0.47$ ). H3 receives directional but not statistical support: the five delivery-team items average 3.18 and the six outcome items 2.88, a difference of 0.30 points, but the exact Mann–Whitney test gives  $U = 23$ ,  $P = .18$  (two-sided). Because the reconstruction cannot pair items within respondents, the participation gradient across stages could not be tested formally.

The management mean (IKM 74.61, 95% CI 67.50–81.72; Table 4) does not differ significantly from 75.00 ( $t = -0.13$ ,  $P = .90$ ) or from 76.61 ( $t = -0.67$ ,  $P = .53$ ), so H4 is not supported. Across all 17 items, the mean of 72.79 (95% CI 68.77–76.82) lies above the 65.00 floor of grade C (adjusted  $P = .009$ ) and does not differ significantly from 75.00 ( $P = .26$ ) or, after correction, from 76.61 (adjusted  $P = .31$ ). Against the red ginger benchmark of 84.75, the six charted overall items and the 17-item mean fall significantly short (adjusted  $P = .009$  and  $P < .001$ ); the management items show the same direction but lose significance after correction ( $P = .01$ , adjusted  $P = .11$ ).

**Table 4. Benchmark tests of mean satisfaction, treating items as units**

| Item group <sup>a</sup>   | <i>k</i> | IKM (95% CI)        | Benchmark                  | <i>t</i> <sup>b</sup> | <i>P</i> | Holm-adjusted <i>P</i> <sup>c</sup> |
|---------------------------|----------|---------------------|----------------------------|-----------------------|----------|-------------------------------------|
| Management                | 8        | 74.61 (67.50–81.72) | 75.00 (report cut-off)     | -0.13                 | .90      | >.99                                |
| Management                | 8        | 74.61 (67.50–81.72) | 76.61 (2017 grade B)       | -0.67                 | .53      | >.99                                |
| Management                | 8        | 74.61 (67.50–81.72) | 65.00 (2017 grade C floor) | 3.19                  | .02      | .11                                 |
| Management                | 8        | 74.61 (67.50–81.72) | 84.75 (red ginger)         | -3.37                 | .01      | .11                                 |
| Overall (6 charted items) | 6        | 71.88 (67.24–76.51) | 75.00 (report cut-off)     | -1.73                 | .14      | .58                                 |
| Overall (6 charted items) | 6        | 71.88 (67.24–76.51) | 76.61 (2017 grade B)       | -2.62                 | .047     | .28                                 |
| Overall (6 charted items) | 6        | 71.88 (67.24–76.51) | 65.00 (2017 grade C floor) | 3.81                  | .01      | .11                                 |
| Overall (6 charted items) | 6        | 71.88 (67.24–76.51) | 84.75 (red ginger)         | -7.14                 | <.001    | .009                                |
| All 17 items              | 17       | 72.79 (68.77–76.82) | 75.00 (report cut-off)     | -1.16                 | .26      | .79                                 |
| All 17 items              | 17       | 72.79 (68.77–76.82) | 76.61 (2017 grade B)       | -2.01                 | .06      | .31                                 |
| All 17 items              | 17       | 72.79 (68.77–76.82) | 65.00 (2017 grade C floor) | 4.10                  | <.001    | .009                                |
| All 17 items              | 17       | 72.79 (68.77–76.82) | 84.75 (red ginger)         | -6.29                 | <.001    | <.001                               |

<sup>a</sup>Items are the units of analysis; IKM =  $25 \times$  mean rating.

<sup>b</sup>One-sample *t* tests with  $k - 1$  degrees of freedom.

<sup>c</sup>Holm adjustment over 12 tests.

## Reconciliation of the Reported Indices

Two reconciliation checks assess the precision stated in the report. For the overall dimension, the six charted items sum to 17.25. If the dimension contains 7 to 10 equally weighted questions, the unreported questions must average between 3.50 and 3.10 to produce the reported 2.965, whereas the charted items alone give 2.875 (IKM 71.9). For delivery, if four equally weighted elements form the index, the unreported reliability item must average 3.23, which corresponds to 51.7 rating points among 16 respondents and is impossible with whole ratings. The question count, the weighting, or the rounding in the report must therefore differ from its description, and three-decimal precision is not justified.

## Action Tiers

Three items fall in tier A (competence, responsibility, responsiveness) and five in tier B (method, fairness, monitoring participation, synergy, closeness) (Table 5, Fig. 2). Three items fall in tier C (relevance, knowledge, and income), each with six or seven respondents rating 2 and an interval extending above 3.00. Four items fall in tier D (skills, environment, implementation participation, adaptability) and two in tier E (continuity, planning participation). Raising the five tier-B items to 3.00 would require 33 one-step improvements among their 80 ratings (41%), and the three tier-C items would require 12 among 48 (25%). The tier A and B assignments do not depend on the one-third convention; lowering the cut-off to 5 of 16 would add skills to tier C, and raising it to 7 would move knowledge to tier D.

Table 5 Action tiers, items, and the improvement required to reach the threshold

| Tier              | Rule                                                  | Items <sup>a</sup>                                                                              | Improvements to reach 3.00 <sup>b</sup> | Management response and next-round target <sup>c</sup>                                                                       |
|-------------------|-------------------------------------------------------|-------------------------------------------------------------------------------------------------|-----------------------------------------|------------------------------------------------------------------------------------------------------------------------------|
| <b>A Sustain</b>  | Lower 95% bound > 3.00                                | Competence, responsibility, responsiveness                                                      | None                                    | Document the CDO approach and formalize peer mentoring. Target: lower bound remains above 3.00                               |
| <b>B Correct</b>  | Upper 95% bound < 3.00                                | Method <sup>d</sup> , fairness <sup>d</sup> , M&E participation, stakeholder synergy, closeness | 33 of 80 ratings (8, 8, 6, 6, 5)        | Redesign and publish targeting rules; establish a joint monitoring team; broaden stakeholders. Target: each mean $\geq 3.00$ |
| <b>C Verify</b>   | CI contains 3.00; $\geq 6$ of 16 rated 2 <sup>e</sup> | Relevance, knowledge, income                                                                    | 12 of 48 ratings (5, 3, 4)              | Record hives, harvest, sales, and household income; develop markets. Target: records for all members                         |
| <b>D Monitor</b>  | CI contains 3.00; 1–5 rated 2                         | Skills, environment, implementation participation, adaptability                                 | 4 of 64 ratings (1, 2, 0, 1)            | Track in the next round. Target: no item falls below 2.75                                                                    |
| <b>E Maintain</b> | CI contains 3.00; none rated 2                        | Continuity, planning participation                                                              | None                                    | Maintain current practice. Target: no respondent rates 2                                                                     |

<sup>a</sup>CDO = Community Development Officer; M&E = monitoring and evaluation.

<sup>b</sup>One-step rating increases among the 16 respondents needed to raise the item mean to 3.00 (Equation 4); figures in parentheses follow the order of the items listed.

<sup>c</sup>Targets are proposals for the next survey, not results.

<sup>d</sup>Assumption-dependent items (Section 2.4).

<sup>e</sup>The one-third cut-off is a convention rather than a statistical result.

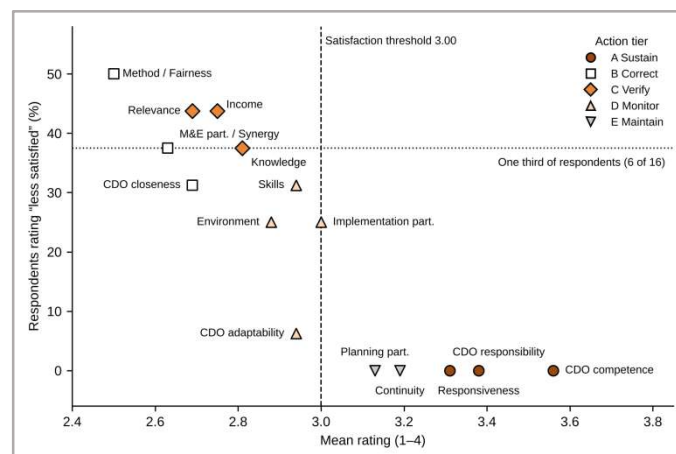


Fig. 2. Action map: mean rating against the share of respondents rating an item “less satisfied”

Tiers follow the rule in Section 2.6. Identical points are labeled together. N = 16. CDO = Community Development Officer; M&E = monitoring and evaluation

The open-ended responses are consistent with this pattern. Respondents attributed income gains to sales of kelulut honey, reported that their new knowledge and skills concern kelulut cultivation, and stated that they share this knowledge with relatives and neighbors. They noted that the company involved the group in designing activities and visited regularly. Every respondent stated that continued company assistance is still needed. The responses provide no explanation of why 44% judged the program less relevant or why half rated method and fairness below satisfactory.

### Delivery Team Versus Program Design

The results indicate that beneficiaries regard the GEMA MADU delivery team favorably, whereas the effectiveness of the program design has not yet been demonstrated. Facilitator competence is the only strength that remains significant after correction for multiple testing, method appropriateness and fairness received the lowest scores, and all four benefit items are statistically indistinguishable from 3.00. The advantage of the delivery team over outcomes (0.30 points) is consistent with expectation–disconfirmation theory, under which good service earns favorable ratings sooner than outcomes that depend on markets and seasons [18,20], although the sample cannot confirm the difference.

This pattern differs from that reported in the public-service IKM literature. In those studies, the weakest elements are process attributes such as speed, schedule certainty, and facilities, cost scores highest in two studies [26,27], and staff performance is generally rated good [25,27]. In GEMA MADU, staff conduct also ranks highest, but the weaknesses are design attributes: whether the method suits beneficiaries, whether benefits are distributed fairly, and whether beneficiaries take part in oversight. An evaluation framework derived from public services, which focuses on waiting times and delays, would therefore fail to detect these weaknesses.

The two lowest items, and the only weaknesses that remain significant after correction, depend on the reconstruction assumption. The report supports only that 50% of respondents judged the method suitable and that 50% judged targeting and distribution less appropriate. If the remaining respondents had all rated fairness 4, the mean would rise to 3.00 with a wide interval (2.45–3.55), and the item would become indistinguishable from 3.00, although never a strength. With eight respondents still rating it 2, it would then move to tier C rather than leave the action list. The robust finding is that half of the respondents rated fairness below satisfactory; the statistical label “weakness” rests on the assumption. Plausible causes include uneven prior experience with bees and perceived inequality in the allocation of hives or training.

### Income, Participation, and Dependency

Income (2.75) and relevance (2.69) have the lowest outcome means, yet the tests cannot classify them as weaknesses, because their effects are small to medium ( $d_z = -0.32$  and  $-0.44$ ) and the sample detects only  $|d_z| \geq 0.75$ . Nevertheless, these values are consistent with the economics of stingless beekeeping. Yields are small, and approximately USD 9 per colony over eight months [14,15] is meaningful for households earning less than IDR 1,000,000 per month but is unlikely to improve livelihoods substantially without scale, diversification, and market access. Satisfaction data alone cannot resolve the status of these items, and the next survey round should be complemented by production and sales records.

Participation declines from planning (3.13) through implementation (3.00) to monitoring and evaluation (2.63). In Arnstein’s terms, the company consults and involves beneficiaries at an early stage, but their role diminishes at the oversight stage [41,42]. This gradient is important because empowerment was the weakest outcome in comparable CSR evidence [8], and community-driven programs require influence beyond planning [43]. Weak stakeholder synergy (2.63) points to a related gap, because kelulut programs need extension services, village government, cooperatives, and buyers that a company cannot readily provide on its own [9,11]. Because contact quality shapes trust in mining communities [7], local peer mentors and messaging groups [17] warrant piloting. All three weaknesses are nominal and require confirmation in a larger sample.

Continuity scored 3.19, with 81% of respondents choosing “continuous” and 19% “highly continuous”, yet every respondent still requires company assistance. These findings are not contradictory. Continuity refers to linked activities, whereas self-reliance refers to operating without the company. Without a planned exit, continuity may foster dependence; graduation milestones such as group-managed hive allocation, independent sales channels, and local peer mentors would address this risk.

### Scale Choice and Comparison with Previous Studies

The interpretation depends more on the choice of scale than on the data. A score of 74.1 is “satisfied” on the report scale but “Fair”, the third of four grades, under the 2017 guideline, and the management mean does not differ significantly from either 75.00 or 76.61. The comparison with other Indonesian IKM studies is indicative because scales and cut-offs differ (Table 6). The GEMA MADU score of 74.1 falls below the 84.75 reported for red ginger cultivation [28], the 83.46 for the Central Lombok civil registry [27], and the 3.48 of 4 (approximately 87 on the converted scale) in Kertajaya [29]. It lies close to the 70.4% of maximum reported for a mining company’s program in Gunung Kidul [30]. None of the studies compared reports confidence intervals or hypothesis tests for the index itself.

**Table 6. Comparison with previous IKM studies in Indonesia**

| Study                    | Setting                                        | $n^a$ | Elements <sup>a</sup> | Index (grade) <sup>b</sup> | Statistical support                              |
|--------------------------|------------------------------------------------|-------|-----------------------|----------------------------|--------------------------------------------------|
| Kartikaningdyah [25]     | Licensing office, Tanjungpinang                | 150   | 14                    | 66.49 (B)                  | None on the index                                |
| Subhan [24]              | Public service mall, Banjarbaru                | 150   | 9                     | 78.8 (B)                   | Validity; $\alpha = 0.718$                       |
| Dinata and Utami [26]    | e-KTP service, Madiun                          | 100   | 9                     | 80.63–80.74 (B)            | None                                             |
| Agus and Ardiyansah [27] | Civil registry, Central Lombok                 | 400   | 14                    | 83.46 (B)                  | None                                             |
| Yuliani et al. [28]      | CSR red ginger program, Banjarbaru             | n/s   | 10                    | 84.75 (very good)          | None                                             |
| Widodo et al. [29]       | CSR programs, Kertajaya                        | 18    | 8 programs            | 3.48 of 4 ( $\approx 87$ ) | None                                             |
| Pebriyani et al. [30]    | Mining company program, Gunung Kidul           | 135   | n/s                   | 3.52 of 5 (70.4%)          | SEM: satisfaction → corporate image              |
| This study               | Kelulut CSR program, coal mining, North Barito | 16    | 17 items              | 74.1 (2004: B; 2017: C)    | Item tests with Holm correction; benchmark tests |

<sup>a</sup>n/s = not stated in the source.

<sup>b</sup>Scales, elements, and grade schemes differ across studies, so the comparison is indicative; grades follow the guideline cited in each study.

Three reporting practices are therefore recommended. First, the interpretation scheme should be fixed before the survey, and the national grade should be reported alongside any custom label. Second, item-level estimates with intervals and corrected tests should replace three-decimal indices. Third, the question counts, weights, and rating counts behind each dimension should be published so that readers can reproduce the index, which the present reconciliation shows is currently impossible. The equal-weights rule also deserves scrutiny, and importance–performance analysis could provide beneficiary-based weights [44].

## Implications for Program Management

Each tier implies a distinct management response. For tier A, the company should document the CDO's approach and formalize it as peer mentoring. For tier B, it should redesign and publish the targeting and distribution rules, form a joint monitoring team with beneficiaries and local extension services, and broaden the stakeholder base. For tier C, it should collect hive, harvest, sales, and household income records and develop markets through quality testing, branding, packaging, and diversification into pot-pollen and propolis [16]. Tier D items should be tracked in the next round, and tier E practices, which support early-stage participation, should be maintained. Finally, the company should adopt a staged exit with milestones for group independence.

## Limitations

The sample is small and purposive, and face-to-face interviews may induce social desirability bias among beneficiaries who depend on company support [45]. The reconstruction is exact for fully charted items but assumption-dependent for method and fairness, and the overall and delivery indices cannot be reproduced exactly. This analysis used aggregated data only; respondent-level data would allow paired tests and reliability coefficients. The survey measures perceptions rather than income, honey production, or colony survival, and the one-third rule for tier C is a convention that other cut-offs would change. Without a baseline or comparison group, changes cannot be attributed to the program, and the results do not extend beyond these 16 beneficiaries.

## 4. CONCLUSION

This study evaluated the satisfaction of 16 beneficiaries of GEMA MADU, a stingless bee honey program of a coal mining company in North Barito. The reported indices (overall 74.116, management 74.609, delivery 72.545) fall in the "satisfied" band, but they are best interpreted as a range of 72.5–74.6 with wide uncertainty that is graded B under the 2004 guideline and C under the 2017 guideline. The tests confirm one robust strength (facilitator competence), two assumption-dependent weaknesses (method appropriateness and fairness), and five nominal effects that require replication. The benefit items, including income, cannot be distinguished from the satisfaction threshold. Reaching the "very satisfied" band on the report scale would require only two additional one-step ratings, whereas the 2017 grade "Good" would require eleven.

The program has built a capable delivery team but has not yet demonstrated that this capability translates into dependable income and self-reliance. The company should continue the program while correcting the tier-B design weaknesses, verifying the tier-C outcomes with production and sales records, introducing participatory monitoring, and planning a staged exit. Future IKM surveys should fix the interpretation scheme in advance, report item-level intervals and corrected tests, disclose the computation, and translate the results into action tiers.

## ACKNOWLEDGEMENTS

The authors thank the management of PT Suprabari Mapanindo Mineral for supporting this evaluation, PT SUCOFINDO Semarang Branch for conducting the survey, and the community members of Lemo I Village who took part as respondents. This research received no external funding.

## COMPETING INTERESTS

Authors Efredi Yudianto, Arianto Sasono, Rapinis Mutiara, Muhammad Rum, Abdul Syukur, Bagus M.M, Muhammad Rizky N, Rayhan Rajoalam Putra P, and Nurlaila Khoirunnisa are employees of PT Tuah Turangga Agung or its subsidiary PT Suprabari Mapanindo Mineral, the company that funds and runs the GEMA MADU program and commissioned the survey analyzed in this study. Author Dewi Hajar has declared that no competing interests exist.

## AUTHORS' CONTRIBUTIONS

Efredi Yudianto conceptualized the study, supervised the work, and reviewed and edited the manuscript. Arianto Sasono provided resources and supervised the program evaluation. Rapinis Mutiara contributed to the conceptualization and methodology. Muhammad Rum validated the results and reviewed and edited the manuscript. Abdul Syukur contributed to the methodology and managed project administration. Bagus M.M carried out the investigation and data curation. Muhammad Rizky N carried out the investigation and formal analysis. Rayhan Rajoalam Putra P curated the data and prepared the visualizations. Nurlaila Khoirunnisa carried out the investigation and formal analysis. Dewi Hajar wrote the first draft of the manuscript. As corresponding author, Efredi Yudianto takes responsibility for the integrity of the data and the accuracy of the analysis. All authors read and approved the final manuscript.

## CONSENT

All respondents gave informed consent to take part in the survey and to the use of the aggregated results for program evaluation and publication. No individual-level data are reported in this manuscript.

## ETHICAL APPROVAL

This study is a secondary analysis of anonymized, aggregated results from a routine program monitoring-and-evaluation survey. It involved no new contact with respondents and no identifiable personal data; therefore, ethical approval was not required.

## REFERENCES

1. Republic of Indonesia. Undang-Undang Nomor 40 Tahun 2007 tentang Perseroan Terbatas [Law No. 40 of 2007 on Limited Liability Companies]. Jakarta: Republic of Indonesia; 2007. Indonesian.
2. Prayogo D. Evaluasi program corporate social responsibility dan community development pada industri tambang dan migas. *Makara Hum Behav Stud Asia*. 2011;15(1):43-58. doi: 10.7454/mssh.v15i1.893. Indonesian.
3. Ministry of Energy and Mineral Resources. Peraturan Menteri ESDM Nomor 41 Tahun 2016 tentang Pengembangan dan Pemberdayaan Masyarakat pada Kegiatan Usaha Pertambangan Mineral dan Batubara [Regulation No. 41 of 2016 on community development and empowerment in mineral and coal mining]. Jakarta: Ministry of Energy and Mineral Resources; 2016. Indonesian.
4. Oktavia M, Yusuf M, Saptawan A. The impact implementation program of corporate social responsibility of PT. Kuansing Inti Makmur toward society development around mining area. *Sriwijaya J Environ*. 2018;3(1):37-42. doi: 10.22135/sje.2018.3.1.37-42
5. Prno J, Slocombe DS. Exploring the origins of 'social license to operate' in the mining sector: perspectives from governance and sustainability theories. *Resour Policy*. 2012;37(3):346-57.

6. Carroll AB. The pyramid of corporate social responsibility: toward the moral management of organizational stakeholders. *Bus Horiz*. 1991;34(4):39-48.
7. Moffat K, Zhang A. The paths to social licence to operate: an integrative model explaining community acceptance of mining. *Resour Policy*. 2014;39:61-70.
8. Prayogo D, Hilarius Y. Efektivitas program CSR/CD dalam pengentasan kemiskinan: studi peran perusahaan geotermal di Jawa Barat. *Masyarakat J Sociol*. 2012;17(1). doi: 10.7454/MJS.v17i1.1189. Indonesian.
9. Muhammad S, Arifin S, Syam R, Tamma S, Hans A, Hanami ZA, et al. Corporate social responsibility programs in mining areas: insights from stakeholder groups in Indonesia. *Cogent Soc Sci*. 2024;10(1):2357675.
10. Freeman RE. *Strategic management: a stakeholder approach*. Boston: Pitman; 1984.
11. Que S, Wang L, Awuah-Offei K, Yang W, Jiang H. Corporate social responsibility: understanding the mining stakeholder with a case study. *Sustainability*. 2019;11(8):2407. doi: 10.3390/su11082407
12. Kahono S, Chantawannakul P, Engel MS. Social bees and the current status of beekeeping in Indonesia. In: Chantawannakul P, Williams G, Neumann P, editors. *Asian beekeeping in the 21st century*. Singapore: Springer; 2018. p. 287-306.
13. Cortopassi-Laurino M, Imperatriz-Fonseca VL, Roubik DW, Dollin A, Heard T, Aguilar I, et al. Global meliponiculture: challenges and opportunities. *Apidologie*. 2006;37(2):275-92. doi: 10.1051/apido:2006027
14. Agussalim, Nurliyani, Umami N, Agus A. The honey and propolis production from Indonesian stingless bee: *Tetragonula laeviceps*. *Livest Res Rural Dev*. 2020;32(8):121.
15. Supeno B, Erwan, Agussalim. The production of honey and pot-pollen from stingless bee *Tetragonula clypearis* and their contribution to increase the farmers income in West Lombok, Indonesia. *Livest Res Rural Dev*. 2022;34(5).
16. Djabbar H, Joshi LK, Apsari NC, Irfan M, Raharjo ST. Pemberdayaan masyarakat melalui budidaya lebah kelulut: program CSR PT Pertamina EP-Sangatta Field. *J Pengabdian Penelitian Kepada Masyarakat*. 2021;2(2):279-90. Indonesian.
17. Sofiyulloh M, Permadi DB, Widayati WT, Soraya E. Evaluasi metode 'texting, sharing and mentoring' dalam program pemberdayaan masyarakat berbasis budidaya lebah klanceng (*Trigona laeviceps*) di Gunungkidul, Yogyakarta. *J Wasian*. 2021;8(2):87-102. doi: 10.20886/jwas.v8i2.6714. Indonesian.
18. Oliver RL. A cognitive model of the antecedents and consequences of satisfaction decisions. *J Mark Res*. 1980;17(4):460-9.
19. Parasuraman A, Zeithaml VA, Berry LL. SERVQUAL: a multiple-item scale for measuring consumer perceptions of service quality. *J Retail*. 1988;64(1):12-40.
20. Van Ryzin GG. Expectations, performance, and citizen satisfaction with urban services. *J Policy Anal Manage*. 2004;23(3):433-48.
21. Likert R. A technique for the measurement of attitudes. *Arch Psychol*. 1932;22(140):1-55.
22. Ministry of State Apparatus Empowerment. Keputusan Menteri PAN Nomor KEP/25/M.PAN/2/2004 tentang Pedoman Umum Penyusunan Indeks Kepuasan Masyarakat Unit Pelayanan Instansi Pemerintah [General guidelines for compiling the community satisfaction index of government service units]. Jakarta: Ministry of State Apparatus Empowerment; 2004. Indonesian.
23. Ministry of State Apparatus Empowerment and Bureaucratic Reform. Peraturan Menteri PANRB Nomor 14 Tahun 2017 tentang Pedoman Penyusunan Survei Kepuasan Masyarakat Unit Penyelenggara Pelayanan Publik [Guidelines for community satisfaction surveys of public service units]. Jakarta: Ministry of State Apparatus Empowerment and Bureaucratic Reform; 2017. Indonesian.
24. Subhan R. Indeks kepuasan masyarakat (IKM) pada mal pelayanan publik Kota Banjarbaru tahun 2022. *J Bisnis Pembangunan*. 2023;12(1):87-94. Indonesian.
25. Kartikaningdyah E. Analisis indeks kepuasan masyarakat terhadap pelayanan publik pada BP2T Kota Tanjungpinang. *J Integrasi*. 2012;4(2):136-47. Indonesian.
26. Dinata NA, Utami DA. Analisis kepuasan masyarakat terhadap pelayanan e-KTP berdasarkan IKM di Kantor Kecamatan Geger Kabupaten Madiun. *Inovant*. 2023;1(2):312-26. Indonesian.
27. Agus, Ardiyansah. Community satisfaction level toward public services at the Population and Civil Registration Office of Central Lombok Regency in 2025: analysis of the community satisfaction index and improvement recommendations. *J Public Policy Gov Local Gov*. 2026;1(1):57-65.
28. Yuliani R, Mubarak Z, Arifah, Norlena. Indeks kepuasan masyarakat terhadap program corporate social responsibility (CSR) Bujur Berstamina: budidaya jahe merah. *JPP J Adm Publik Pembangunan*. 2023;5(1):14-21. Indonesian.
29. Widodo WE, Cholidah SN, Isnaeni AP, Wibowo KT, Abriandi E. Mengukur kepuasan masyarakat pada program CSR di Desa Kertajaya: sebuah analisis compass sustainability method. *J Pemberdayaan Masy*. 2019;3(1):31-54. doi: 10.14421/jpm.2019.031-02. Indonesian.
30. Pebriyani M, Adnyano AAIA, Rande SA. Kajian program PPM community development (COMDEV) di PT. Sugih Alamanugroho di Desa Bedoyo, Kecamatan Ponjong, Kabupaten Gunung Kidul, Daerah Istimewa Yogyakarta. In: *Prosiding Nasional Rekayasa Teknologi Industri dan Informasi XVII (ReTII)*; 2022. p. 182-8. Indonesian.
31. Bunga YR, Seran MY, Mappiasse, Putera MI. Persepsi masyarakat terhadap program pemberdayaan masyarakat binaan Balai Taman Nasional Komodo (studi kasus: Kampung Rinca). *JAVOK J Akad Vokasi*. 2024;3(2):14-25. Indonesian.
32. Republic of Indonesia. Undang-Undang Nomor 27 Tahun 2022 tentang Pelindungan Data Pribadi [Law No. 27 of 2022 on Personal Data Protection]. Jakarta: Republic of Indonesia; 2022. Indonesian.
33. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet*. 2007;370(9596):1453-7. doi: 10.1016/S0140-6736(07)61602-X
34. Cohen J. *Statistical power analysis for the behavioral sciences*. 2nd ed. Hillsdale (NJ): Lawrence Erlbaum Associates; 1988.
35. Norman G. Likert scales, levels of measurement and the 'laws' of statistics. *Adv Health Sci Educ Theory Pract*. 2010;15(5):625-32. doi: 10.1007/s10459-010-9222-y
36. Sullivan GM, Artino AR Jr. Analyzing and interpreting data from Likert-type scales. *J Grad Med Educ*. 2013;5(4):541-2.
37. Cumming G. The new statistics: why and how. *Psychol Sci*. 2014;25(1):7-29.
38. Holm S. A simple sequentially rejective multiple test procedure. *Scand J Stat*. 1979;6(2):65-70.
39. Wilson EB. Probable inference, the law of succession, and statistical inference. *J Am Stat Assoc*. 1927;22(158):209-12.

40. Brown LD, Cai TT, DasGupta A. Interval estimation for a binomial proportion. *Stat Sci.* 2001;16(2):101-33. doi: 10.1214/ss/1009213286
41. Arnstein SR. A ladder of citizen participation. *J Am Inst Plann.* 1969; 35(4):216-24.
42. Cornwall A. Unpacking 'participation': models, meanings and practices. *Community Dev J.* 2008;43(3):269-83.
43. Mansuri G, Rao V. Community-based and -driven development: a critical review. *World Bank Res Obs.* 2004; 19(1):1-39.
44. Martilla JA, James JC. Importance-performance analysis. *J Mark.* 1977; 41(1):77-9.
45. Nederhof AJ. Methods of coping with social desirability bias: a review. *Eur J Soc Psychol.* 1985;15(3):263-80. doi: 10.1002/ejsp.2420150303

## DEFINITIONS, ACRONYMS, ABBREVIATIONS

**CDO:** Community Development Officer

**CI:** Confidence interval

**CSR:** Corporate social responsibility

**GEMA MADU:** Gerakan Masyarakat Budidaya Madu Kelulut (Stingless Bee Honey Cultivation Movement)

**IKM:** Indeks Kepuasan Masyarakat (Community Satisfaction Index)

**M&E:** Monitoring and evaluation

**SMM:** PT Suprabari Mapanindo Mineral

\*\*\*\*\*