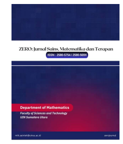
Zero : Jurnal Sains, Matematika, dan Terapan
E-ISSN : 2580-5754; P-ISSN : 2580-569X
Volume 10, Number 2, 2026
DOI: 10.30829/zero.v10i2.29707
Page: 574-587



Development of a Tourism Awareness Index and its Determinants among Tourism Workers

[1]Sarni Maniar Berliana
Research Unit of Sustainable Development Goals, Politeknik Statistika STIS, Jakarta, 13330, Indonesia

[2]Krismanti Tri Wahyuni
Research Unit of Sustainable Development Goals, Politeknik Statistika STIS, Jakarta, 13330, Indonesia

[3]Victoria P. Valenzuela
CSSP-DPAG-Bulacan State University, Bulacan, 3000, Philippines

[4]Joshua G. Meneses
San Miguel National High School, San Juan, San Miguel, Bulacan, 3011, Philippines

[5]Wahyuni Andriana Sofa
Research Unit of Economy and Policy, Politeknik Statistika STIS, Jakarta, 13330, Indonesia

[6]Maria A. Hasiholan Siallagan
BPS-Statistics Indonesia, Jl. Dr. Sutomo 6-8, Jakarta, 10710, Indonesia

[7]Ditto Satrio Wicaksono
BPS-Statistics of Bulukumba Regency, Jl. Ahmad Yani No.18, Bulukumba, 92517, Indonesia

Article Info

Article history:

Accepted: 10 July 2026

Keywords:

Composite Index;
Factor Analysis;
Ordinal Logistic Regression;
Sapta Pesona;
Tourism Awareness;
Tourism Workers.

ABSTRACT

Tourism worker awareness is essential for service quality and sustainable tourism development, yet a validated composite measure remains lacking. This study develops the Tourism Awareness Index (TAI) based on the Sapta Pesona framework using survey data from 868 tourism workers in Malang Regency, Indonesia. Six dimensions showed satisfactory internal consistency (Cronbach's $\alpha = 0.65–0.81$), while one dimension had α slightly above 0.60. Factor analysis with principal component extraction identified a three-factor structure explaining 74.2% of total variance, supported by a KMO value of 0.91 and a significant Bartlett's test. Ordered logistic regression revealed that workers aged 45 years and above ($OR = 1.42$, $p = 0.042$) and those with at least a high school education ($OR = 1.60$, $p = 0.004$) were more likely to achieve higher TAI levels, while lack of foreign language proficiency reduced this likelihood ($OR = 0.57$, $p = 0.003$). The TAI offers a statistically validated tool and may support workforce development and tourism policy.

Corresponding Author:

Sarni Maniar Berliana,
Research Unit of Sustainable Development Goals,
Politeknik Statistika STIS, Jakarta, Indonesia
Email: sarni@stis.ac.id

1. INTRODUCTION

Tourism is one of the fastest-growing global industries, contributing significantly to economic growth, employment, and cultural exchange. In 2015, tourism accounted for 10% of global GDP and one in ten jobs worldwide. However, the COVID-19 pandemic disrupted this trajectory [1], particularly in destinations such as Bali, Indonesia [2]. These challenges underscored the need for sustainable tourism strategies that balance economic, social, and environmental dimensions [3], [4].

Several indices have been developed to evaluate tourism performance and competitiveness. The World Economic Forum's Travel & Tourism Development Index (TTDI) benchmarks enabling environments, policy conditions, infrastructure, and sustainability across 140 economies [5]. At the national level, Indonesia's National Tourism Development Index (IPKN) assesses local government contributions through six dimensions, including governance, destination, accessibility, amenities, local economy, and socio-environmental sustainability [6], [7]. Other studies have constructed composite indices following OECD guidelines [8], [9] that focused on efficiency in Mediterranean destinations [10], heritage site competitiveness in Egypt [11], touristification in Greek islands [12], and vulnerability of Small Island Developing States [13], as well as sustainability-focused indices using other methods [14], [15], [16]. While these indices demonstrate methodological diversity, they primarily address destination-level competitiveness and sustainability rather than individual worker awareness.

Recent studies emphasize that workforce skills, education, and local context are critical for tourism sustainability [17], [18], [19]. However, no research has yet developed a composite index specifically to measure tourism awareness among tourism workers. To address this gap, the present study proposes the Tourism Awareness Index (TAI) grounded in the Sapta Pesona framework [20], a nationally endorsed program under Regulation No. PM.04/UM.001/MKP/2008. Sapta Pesona consists of seven complementary dimensions—safety, orderliness, cleanliness, coolness, beauty, politeness, and memorability—that are theoretically coherent and suitable for index construction.

Accordingly, this study aims to develop the TAI based on the Sapta Pesona framework and to analyze the factors influencing worker awareness levels using ordinal logistic regression [21]. The study hypothesizes that worker awareness, as measured by the TAI, is significantly influenced by sociodemographic characteristics (age, gender, education, training participation, and foreign language ability) and residential factors (type of area and subdistrict). This research contributes methodologically by advancing composite index construction in tourism studies and substantively by providing empirical evidence to support workforce development and sustainable tourism policy.

2. RESEARCH METHOD

2.1 Research Design

This study employed a quantitative design that combined composite index construction with inferential statistical modeling. The Tourism Awareness Index (TAI) was developed to measure tourism awareness among workers, while ordinal logistic regression was used to identify its determinants. This approach was chosen to ensure both measurement robustness and explanatory power.

The TAI was constructed based on the Sapta Pesona framework stipulated in the Ministry of Culture and Tourism Regulation Number 4 Year 2008 [20]. This framework consists of seven dimensions—safety, orderliness, cleanliness, coolness, beauty, politeness, and memorability—which serve as foundational elements for tourism development in Indonesia. These dimensions reflect essential aspects of tourist experience, ranging from basic security and hygiene to aesthetic appeal, hospitality, and lasting impressions. A total of 38 items were developed to represent the seven dimensions, measured on a 5-point Likert scale (1 = strongly disagree to 5 = strongly agree). Higher scores indicate greater alignment with Sapta Pesona principles and thus higher tourism awareness. This operationalization grounds the TAI in a nationally recognized framework while ensuring it is measurable and replicable.

2.2 Data

Malang Regency was selected as the study area due to its significant tourism potential and leading contribution to East Java's tourism sector, recording 1.5 million domestic tourist visits in 2021 [22]. Data were collected through an enumeration survey of tourism workers between January 23 and February 3, 2023, using a three-stage cluster sampling design. Three subdistricts (Pujon, Poncokusumo, and Bantur) were selected using probability proportional to size (PPS). In total, 129 census blocks and 712 households were surveyed, yielding 868 eligible respondents (119.1% of the minimum required sample size of 598 households calculated via the Multiple Indicator Cluster Survey (MICS) method).

The questionnaire was pilot-tested on 57 respondents in Pujon to assess clarity, consistency, and feasibility, and subsequently revised. To minimize enumerator bias, 129 trained enumerators from Politeknik Statistika STIS received standardized training, and field supervision was implemented throughout the data collection process.

2.3 Ethical Considerations

This study adhered to institutional and national ethical guidelines for social research. Participation was voluntary, and informed consent was obtained from all respondents. Confidentiality of responses was strictly maintained, and data were used solely for academic purposes. The involvement of trained enumerators further ensured ethical compliance during the survey.

2.4 Composite Index Construction

This study followed the ten steps of composite index construction recommended by the OECD [9]. These include: (1) developing a theoretical framework, (2) selecting variables, (3) imputing missing data, (4) performing multivariate analysis, (5) normalizing data, (6) applying weighting and aggregation, (7) conducting robustness and sensitivity checks, (8) revisiting the details, (9) linking to other variables, and (10) presenting and disseminating the results. In this study, the first two steps were addressed by adopting the Sapta Pesona concepts, while the third step (imputation of missing data) was omitted as the analysis used only complete observations.

2.4.1 Instrument Validity and Reliability

Before constructing the composite index, instrument validity and reliability testing were conducted to ensure the questionnaire accurately and consistently measured tourism awareness. Validity refers to the extent to which the instrument measures what it is intended to measure, while reliability concerns the consistency and stability of the results [23], [24].

a. Validity Test

Construct validity was assessed through internal consistency by examining the correlation between each item and the total score using Pearson's correlation coefficient [25]. An item-total correlation of 0.3 or higher was considered acceptable, indicating that the item contributed meaningfully to the overall construct. Items with low correlations were reviewed or removed to improve the scale's validity [26].

b. Reliability Test

Internal consistency reliability was evaluated using Cronbach's alpha, which measures the degree to which items in a scale are correlated with one another [27]. A Cronbach's alpha value above 0.7 was considered indicative of good internal consistency, ensuring that the items reliably captured the intended construct of tourism awareness.

2.4.2 Factor Analysis

Factor analysis was conducted in step 4 to explain the variability among observed variables in terms of a smaller number of unobserved factors [28]. The factor model can be expressed as

$$(\mathbf{Y}-\boldsymbol{\mu})_{q\times 1} = \mathbf{L}_{q\times m}\mathbf{F}_{m\times 1} + \boldsymbol{\varepsilon}_{q\times 1} \tag{1}$$

where $\mathbf{Y}$ denotes observable random vector with q components, having mean $\boldsymbol{\mu}$ and a covariance matrix $\boldsymbol{\Sigma}$. $\mathbf{L}$ represents the matrix of factor loadings, while $\mathbf{F}$ (common factors) and $\boldsymbol{\varepsilon}$ (errors or specific factors) are both unobservable random vectors.

Prior to analysis, data suitability was assessed using the Kaiser-Meyer-Olkin (KMO) measure, Measure of Sampling Adequacy (MSA), and Bartlett's Test of Sphericity. KMO values above 0.6, MSA values greater than 0.5, and a significant Bartlett's test ($p < 0.05$) indicate that the data are suitable for factor analysis [29], [30], [31].

In this study, factor analysis was used to reduce the seven Sapta Pesona dimension scores into a smaller set of underlying factors and to determine appropriate weights for the TAI. Although often referred to broadly as factor analysis, this study specifically employed principal component as the extraction method, chosen because the goal was to construct a composite index by maximizing explained variance [32].

a. Factor Extraction and Rotation

The number of factors was determined based on the cumulative proportion of explained variance. Factors were retained until the cumulative variance reached approximately 80–90% [28]. In social science research, a cumulative variance of 70–80% is often considered acceptable if it produces a clear and parsimonious structure [32], [33]. Rotation was applied to improve interpretability. Three rotation methods were tested: Varimax (orthogonal), Promax (oblique), and Oblimin (oblique). Varimax assumes uncorrelated factors, while Promax and Oblimin allow factor correlations, which is often more realistic in social science data [34], [35], [36].

b. Factor Scores and Weighting

Factor scores were estimated using the regression method [28]

$$\hat{\mathbf{f}}_k = \hat{\mathbf{L}}'\mathbf{S}^{-1}(\mathbf{y}_k - \bar{\mathbf{y}}) \tag{2}$$

where $\hat{\mathbf{f}}_k$ denotes the factor score, $\mathbf{S}$ represents the covariance matrix, k is the number of extracted factors $(k = 1, \cdots, m)$, and $\hat{\mathbf{L}}$ is the estimated loading matrix.

Factor weights were derived from the squared loadings

$$w_k = \frac{l_k^2}{\sum_{k=1}^{m} l_k^2} \tag{3}$$

where w_k is the weight for the k^{th} factor, l_k is the k^{th} factor loading, and $k = 1, \dots, m$ [28].

2.4.3 Normalization

Although the OECD recommends normalization before weighting [9], this study applied weighting first because the dimension scores were already on comparable scales. After aggregation, the composite index was rescaled to a 0–100 range using the formula

$$TAS_{norm} = \frac{TAS - TAS_{min}}{TAS_{max} - TAS_{min}} \times 100 \tag{4}$$

where TAS is the original tourism awareness score, TAS_{min} and TAS_{max} are the minimum and maximum score of the tourism awareness, and TAS_{norm} is the rescaled score ranging from 0 to 100. This rescaling step improves interpretability and comparability [9].

2.4.4 Weighting & Aggregation

Weighting reflects the relative importance of each factor in the composite index. This study used data-driven weighting based on principal component extraction. Aggregation was performed using a weighted linear sum [9]

$$TAI = \frac{1}{m}(w_1 TAS_{norm}^1 + w_2 TAS_{norm}^2 + \cdots + w_m TAS_{norm}^m) \tag{5}$$

where TAI is the composite index of tourism awareness, w_k is the weight for the k^{th} factor, and TAS_{norm}^k is the normalized score of tourism awareness of the k^{th} factor.

2.4.5 Robustness Assessment

To evaluate the stability of the TAI, nine scenarios were constructed by combining three principal component extraction approaches with three rotation techniques (Varimax, Promax, and Oblimin). Robustness was assessed using Spearman's rank correlation and mean absolute rank differences (ranking shifts) [37], [38]. A high correlation indicates that the relative positions of observations remain stable despite changes in construction methods [38], [39]. Smaller shifts suggest that the index is resilient to methodological uncertainty [40]. These two metrics are widely recommended because they are distribution-free, focus on ordinal information (which is central to index interpretation), and allow for transparent evaluation of index stability [9]. The formulas used are [9], [41]

$$\rho_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2 - 1)} \tag{6}$$

$$\bar{R}_\iota = \frac{1}{n}\sum_{i=1}^{n} \left| Rank_{ref}(TAI_i) - Rank(TAI_i) \right| \tag{7}$$

where ρ_s is Spearman's correlation coefficient, $\bar{R}_\iota$ is average ranking shifts, n is the number of observations, and $d_i^2 = \sum_{i=1}^{n}\left[Rank_{ref}(TAI_i) - Rank(TAI_i)\right]^2$. The reference ranking for the TAI analysis is the original rank given to the observation by the deterministic version of the composite indicator.

2.4.6 Revisiting the Details

This stage involves reviewing the dimensional structure and data consistency to ensure coherence and interpretability. In this study, the TAI was revisited by examining score distributions across dimensions and socio-demographic characteristics. Results were presented in tabular form, disaggregated by dimension and demographic profile, thereby enhancing the index's interpretability and policy relevance.

2.4.7 Linking to Other Variables

Although the OECD recommends linking composite indices to other variables for external validation [9], this step was not performed. As the TAI is a novel construct with no comparable individual-level indices available, additional linkages risked spurious associations. This deliberate omission preserves the integrity and clarity of the TAI framework.

2.5 Ordinal Logistic Regression

Ordinal logistic regression is a statistical technique used when the response variable is ordinal—that is, it consists of categories with a natural order but unknown or unequal distances between them. Examples include satisfaction levels ("Very Dissatisfied," "Dissatisfied," "Neutral," "Satisfied," "Very Satisfied") or educational attainment. Unlike multinomial logistic regression, which is used for unordered categorical outcomes, ordinal logistic regression explicitly accounts for the ordered nature of the response variable [42], [43].

The most widely used specification is the proportional odds model (POM), introduced by [44]. For a response variable with J ordered categories $(\mathbf{Y}^* = 1,2,\dots,J)$ and predictors $X_1,\dots,X_p$, the cumulative logit model is written as

$$\log\left(\frac{P(\mathbf{Y}^* \leq j \mid \mathbf{X})}{P(\mathbf{Y}^* > j \mid \mathbf{X})}\right) = \theta_j - \mathbf{X}'\boldsymbol{\beta}, \quad j = 1,2,\dots,J-1 \tag{8}$$

where θ_j is the intercept specific to threshold j, and $\mathbf{X}'\boldsymbol{\beta}$ are linear combination of predictors. This parallel lines assumption implies that the effect of each predictor is the same regardless of the cut-off point.

2.5.1 Test of Parallel Lines Assumption

A critical assumption of the POM is that regression coefficients are equal across all cumulative logits. This is known as the parallel lines or proportional odds assumption. This assumption was tested using the Brant test [45]. The hypotheses are

$$\begin{gathered} H_0: \beta_j = \beta \quad \text{for all } j = 1,\dots,J-1 \text{ (parallel lines assumption holds)} \\ H_1: \beta_j \neq \beta \quad \text{for some } j \text{ (parallel lines assumption violated)} \end{gathered} \tag{9}$$

The Brant test compares estimated coefficients across logits using a chi-square statistic

$$\chi^2 = \left(\mathbf{D}\widehat{\boldsymbol{\beta}}\right)^{\mathrm{T}} \left[\mathbf{D} \times \widehat{\mathbf{Var}}\left(\widehat{\boldsymbol{\beta}}\right) \times \mathbf{D}^{\mathbf{T}}\right]^{-1} \left(\mathbf{D}\widehat{\boldsymbol{\beta}}\right) \sim \chi^2_{(J-2)p} \tag{10}$$

where $\mathbf{D}$ is a contrast matrix, $\widehat{\boldsymbol{\beta}}$ is the vector of estimated coefficients, and $\widehat{\mathbf{Var}}\left(\widehat{\boldsymbol{\beta}}\right)$ is the estimated covariance matrix. Rejection of the null hypothesis $(p < \alpha)$ indicates a violation of the proportional odds assumption. Such violations are common with large samples, numerous predictors, or continuous variables [21]. In cases where only some predictors violate the assumption, the Partial Proportional Odds Model (PPOM) may be more appropriate [46], as it allows selected predictors to have varying effects across thresholds while maintaining proportionality for others.

2.5.2 Interpretation of Odds Ratios

The interpretation of regression coefficients in the proportional odds ordinal logistic regression model requires careful attention because the model is based on cumulative logits. A positive coefficient $(\beta_k > 0)$ indicates that higher values of the predictor X_k are associated with higher categories of the ordinal outcome variable Y^*. The exponentiated coefficient, $\exp(\beta_k)$, represents the odds ratio. Because the model estimates cumulative probabilities, there are two logically equivalent but directionally different ways to interpret this odds ratio [47].

One approach states that for a one-unit increase in X_k, the odds of being in a lower or equal category $(Y^* \leq j)$ versus being in a higher category $(Y^* > j)$ are multiplied by $\exp(\beta_k)$, for all $j = 1,\dots,J-1$. It is often more intuitive to reverse the direction of interpretation. An odds ratio of $\exp(\beta_k)$ means that for a one-unit increase in X_k, the odds of being in a higher category $(Y^* > j)$ versus being in a lower or equal category $(Y^* \leq j)$ are multiplied by $\exp(\beta_k)$, for all $j = 1, \dots, J-1$. If $\exp(\beta_k) > 1$, the predictor increases the odds of higher outcome categories; if $\exp(\beta_k) < 1$, it decreases the odds. This second interpretation is generally preferred because it is more intuitive and avoids double negatives.

3. RESULT AND ANALYSIS

The initial stage in constructing the Tourism Awareness Index (TAI) involved harmonizing the direction of negatively worded items so that higher Likert scale values consistently indicated greater tourism awareness. Following recoding, validity and reliability assessments were conducted.

Reliability testing results in Table 1 showed that Cronbach's alpha for six dimensions ranged from 0.65 to 0.81, indicating good internal consistency. One dimension (Orderliness) recorded an alpha of 0.634. Although slightly below the conventional 0.7 threshold, this value is still acceptable for exploratory research, as values above 0.6 are considered usable in early-stage scale development [27], [32], [36]. Thus, all dimensions demonstrated adequate reliability.

Table 1. Results of Validity and Reliability Test

Dimension	Item	Overall Cronbach's Alpha	Cronbach Alpha	Pearson's Correlation
Safety	I provide initial assistance to tourists in case of a medical emergency.	0.810	0.795	0.577
	I help ensure the tourism environment remains safe.		0.787	0.654
	I ensure that public spaces used by tourists are secure for their protection.		0.785	0.671
	I assist in providing essential safety information to visitors before they use tourism facilities.		0.786	0.668
	I consistently offer information to tourists about attractions, including details on locations, accommodations, dining options, and transportation.		0.791	0.622
	I believe there is no need to explain security rules and restrictions in tourist areas.		0.813	0.526
	I cause disruptions at tourist sites.		0.806	0.453
	Before assisting tourists, I ensure I am in good health to prevent the spread of infectious diseases.		0.794	0.580
	I do not concern myself with visitors who ignore health precautions.		0.81	0.507
	I welcome feedback from tourists on how to improve tourism security services.		0.795	0.576
	I treat all tourists equally in the services I provide.		0.798	0.541
	I am always ready to assist tourists who experience accide[32nts.		0.793	0.592
Orderliness	I practice a queuing system into my daily routine.	0.634	0.593	0.656
	I take steps to avoid any harm to property or the environment caused by tourists.		0.637	0.652
	I am punctual in completing my tasks.		0.524	0.736
	I have a consistent attitude and dedication toward my work.		0.500	0.732
Cleanliness	I consistently reprimand tourists who litter.	0.667	0.621	0.620
	I always dispose of trash properly.		0.621	0.597
	I make sure not to pollute tourist facilities and infrastructure.		0.627	0.590
	I disregard tourists who litter at tourist attractions.		0.605	0.683
	I tolerate tourists who smoke irresponsibly.		0.651	0.614
	I make an effort to keep my clothing clean and tidy at work.		0.626	0.587
Coolness	I make sure not to damage plants in tourist areas or along tourist routes.	0.694	0.600	0.745
	I reprimand tourists who harm plants in tourist environments.		0.690	0.704
	I am willing to assist in greening efforts by planting trees.		0.610	0.754
	I enjoy a refreshing and green tourist environment.		0.620	0.704
Beauty	I contribute to preserving the beauty of tourist attractions.	0.653	0.470	0.857
	I assist in creating a pleasant living environment.		0.501	0.867

Dimension	Item	Overall Cronbach's Alpha	Cronbach Alpha	Pearson's Correlation
Politeness	I consistently smile and greet tourists when interacting with them.	0.682	0.508	0.868
	I am always willing to assist tourists who need help, voluntarily.		0.528	0.874
Memorability	I help introduce local arts and culture to tourists.	0.799	0.731	0.811
	I take part in promoting traditional activities of local communities.		0.710	0.837
	I introduce tourists to local food and beverages.		0.775	0.747
	I contribute to promoting souvenirs sold in traditional markets or souvenir centers.		0.773	0.764

Dimension scores were then calculated and subjected to factor analysis. The overall KMO-MSA value was 0.91, the KMO-MSA values for each dimension exceeded 0.5, and Bartlett's Test of Sphericity was significant ($p < 0.05$), confirming the data's suitability for factor analysis. The scree plot (Figure 1) supported retaining three factors, with cumulative variance reaching approximately 77%. The final three-factor solution explained 74.2% of the total variance (31.3%, 28.5%, and 14.4% respectively).

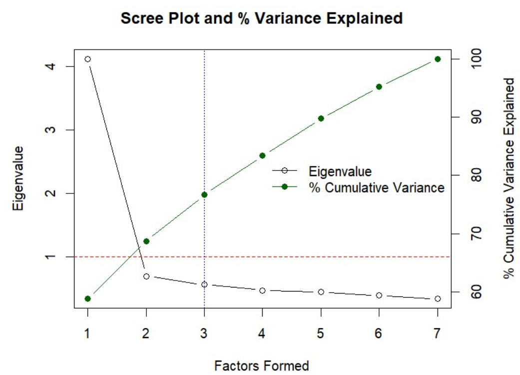


Figure 1. Scree Plot and Cumulative Percentage of Variance Explained

The rotated factor loadings in Table 2 revealed three distinct factors: (1) Safety–Orderliness–Cleanliness, (2) Coolness–Beauty–Politeness, and (3) Memorability. Factor 1 was dominated by Cleanliness (loading = 0.936), followed by Orderliness (0.800) and Safety (0.756). This pattern indicates that Cleanliness is the most dominant indicator within this factor, while Safety and Orderliness serve as closely related supporting components.

Although the third factor consists of a single dimension, Memorability consistently emerged as a distinct factor across all nine scenarios (Table 4) with very high loadings. This reflects its unique conceptual contribution within the Sapta Pesona framework, capturing the experiential and emotional aspects of tourism awareness that are not subsumed under the other two factors. Single-indicator factors are generally viewed as less stable; however, they may be retained when the indicator represents a conceptually unique dimension that cannot be adequately measured by other items, provided the factor loading is sufficiently high [32], [35]. The factor score determinacy coefficient for this factor was 0.9916, which is well above the commonly recommended threshold of 0.80 [48]. This high determinacy indicates that the factor scores are sufficiently stable and reliable for further analysis, despite the factor's single-indicator composition.

Table 2. Reliability Test and Factor Loadings for Each Dimension of Seven Charms

Dimension	Cronbach's α	MSA	Factor Loading		
			Factor 1	Factor 2	Factor 3
Safety	0.810	0.92	**0.756**		
Orderliness	0.634	0.92	**0.800**		
Cleanliness	0.667	0.91	**0.936**	-0.111	
Coolness	0.694	0.90	0.312	**0.604**	
Beauty	0.653	0.90		**0.828**	0.112
Politeness	0.682	0.92		**0.960**	
Memorability	0.799	0.94			**0.989**

Factor scores were weighted according to the proportion of variance explained by each factor (0.42, 0.39, and 0.19) as shown in Table 3 and aggregated into the TAI. The weighting scheme reflects the relative contribution of each factor to the overall construct of tourism awareness. Factor 1, defined by *Safety-Orderliness-Cleanliness*, carries the largest weight (0.42), indicating its dominant role in shaping the TAI. Factor 2 (*Coolness-Beauty-Politeness*) contributes a comparable proportion (0.39), while Factor 3 (*Memorability*) provides a smaller yet meaningful contribution (0.19).

Table 3. Factor Weights

Factor	Weight
1	0.42
2	0.38
3	0.19

To evaluate the robustness of the TAI, nine scenarios were constructed by combining three principal component extraction approaches with three rotation techniques (Varimax, Promax, and Oblimin). Although all scenarios used principal component extraction, they differed in criteria for retaining components (Table 4). The first approach applied principal component to the covariance matrix and retained components with eigenvalues greater than 1.0 (Kaiser's criterion). The second used the correlation matrix and retained components until cumulative explained variance reached 80–90%. The third also used the correlation matrix but applied a more parsimonious threshold of about 75%. Employing multiple extraction criteria with varying variance thresholds allowed a robust assessment of methodological uncertainty, balancing parsimony and explanatory power. Following extraction, three rotation methods were applied: Varimax (orthogonal), Promax (oblique), and Oblimin (oblique). Orthogonal rotation assumes uncorrelated factors, while oblique rotations allow correlation, which is often more realistic in social science research [35].

Table 4. Nine Scenarios of Principal Component Extraction with Different Factor Retention Criteria and Rotation Methods for Constructing the TAI

Scenario	Factor Extraction Method	Rotation Method	Factors Formed	Eigenvalue Proportions (Weights)
1	PC on Covariance Matrix (Eigenvalue > 1)	Oblimin	5 factors	(0.25, 0.24, 0.22, 0.20, 0.09)
2	PC on Correlation Matrix (80–90% variance)	Oblimin	4 factors	(0.33, 0.25, 0.21, 0.20)
3	PC on Correlation Matrix (75% variance)	Oblimin	3 factors	(0.42, 0.38, 0.20)
4	PC on Covariance Matrix (Eigenvalue > 1)	Promax	5 factors	(0.31, 0.21, 0.18, 0.17, 0.13)
5	PC on Correlation Matrix (80–90% variance)	Promax	4 factors	(0.35, 0.26, 0.20, 0.18)
6	PC on Correlation Matrix (75% variance)	Promax	3 factors	(0.42, 0.38, 0.19)
7	PC on Covariance Matrix (Eigenvalue > 1)	Varimax	5 factors	(0.25, 0.19, 0.16, 0.16, 0.15)
8	PC on Correlation Matrix (80–90% variance)	Varimax	4 factors	(0.33, 0.29, 0.20, 0.19)
9	PC on Correlation Matrix (75% variance)	Varimax	3 factors	(0.41, 0.38, 0.20)

To assess the robustness of TAI rankings, Spearman rank correlations were calculated across nine scenarios. As shown in Table 5, correlations were exceptionally high (0.956–0.999), averaging above 0.98. This indicates that the relative ordering of tourism awareness levels is highly stable and minimally affected by factor retention criteria or rotation methods, confirming the robustness of the composite index.

Table 5. Spearman Rank Correlation Matrix of TAI Across Nine Methodological Scenarios

Scenario	1	2	3	4	5	6	7	8	9	Average Correlation
1	1	0.9923	0.9849	0.9655	0.9713	0.9698	0.9670	0.9612	0.9564	0.9743
2	0.9923	1	0.9911	0.9768	0.9775	0.9758	0.9758	0.9668	0.9625	0.9798
3	0.9849	0.9911	1	0.9799	0.9817	0.9836	0.9837	0.9793	0.9698	0.9838
4	0.9655	0.9768	0.9799	1	0.9989	0.9981	0.9963	0.9937	0.9958	0.9894
5	0.9713	0.9775	0.9817	0.9989	1	0.9995	0.9972	0.9959	0.9972	0.9910
6	0.9698	0.9758	0.9836	0.9981	0.9995	1	0.9977	0.9973	0.9971	0.9910
7	0.9670	0.9758	0.9837	0.9963	0.9972	0.9977	1	0.9986	0.9963	0.9903
8	0.9612	0.9668	0.9793	0.9937	0.9959	0.9973	0.9986	1	0.9974	0.9878
9	0.9564	0.9625	0.9698	0.9958	0.9972	0.9971	0.9963	0.9974	1	0.9858

As a complementary measure, the mean absolute rank differences were computed to capture average changes in ranking positions (Table 6). Differences ranged from 4.87 to 51.06, with an overall average of 27. As shown in Tables 5 and 6, all scenarios produced highly consistent rankings and mean absolute rank differences. However, Scenario 6 was ultimately selected because it combined high stability (average rank difference 18.00, comparable to Scenario 5's 18.11) with greater parsimony, reducing the dimensionality to three factors instead of four.

Table 6. Mean Absolute Rank Differences of TAI Across Nine Methodological Scenarios

Scenario	1	2	3	4	5	6	7	8	9	Average Rank Difference
1	0	21.30	29.96	46.05	41.44	42.60	44.42	48.15	51.06	36.11
2	21.30	0	20.93	36.36	36.19	37.52	37.30	44.08	47.47	31.24
3	29.96	20.93	0	33.52	32.65	31.00	30.41	34.57	42.56	28.40
4	46.05	36.36	33.52	0	7.98	9.78	13.80	18.43	15.84	20.20
5	41.44	36.19	32.65	7.98	0	4.87	12.03	14.85	12.97	18.11
6	42.60	37.52	31.00	9.78	4.87	0	10.73	12.17	13.31	**18.00**
7	44.42	37.30	30.41	13.80	12.03	10.73	0	8.86	14.82	19.15
8	48.15	44.08	34.57	18.43	14.85	12.17	8.86	0	12.01	21.46
9	51.06	47.47	42.56	15.84	12.97	13.31	14.82	12.01	0	23.34

After assessing the robustness of the composite index, the study revisited the TAI by examining scores and their dimensions across workers' socio-demographic backgrounds (Table 7). Clear patterns emerged: education, training participation, foreign language ability, and rural residence were associated with higher awareness. Workers with at least a high school education scored higher overall (60.24 vs. 56.89), especially in the Memorable dimension (70.46 vs. 65.86). Training participants also achieved higher scores across all dimensions, notably Memorable (72.12 vs. 67.76). Similarly, workers with foreign language proficiency recorded higher scores (63.57 vs. 57.57), underscoring the role of communication skills. Rural workers outperformed urban counterparts (61.17 vs. 58.11), with a marked difference in Memorable (70.81 vs. 67.79). Differences by gender and subdistrict were modest, while older workers (aged 45 and above) tended to show higher awareness, particularly in the Coolness–Beauty–Politeness dimension.

Table 7. TAI and Dimension Scores by Background Characteristics

Characteristics	TAI	Dimension		
		Safety-Orderliness-Cleanliness	Coolness-Beauty-Politeness	Memorable
Gender				
Male	58.55	67.49	64.66	68.77
Female	59.10	67.31	66.68	67.99
Age				
15-44	58.13	66.98	64.43	68.83
>=45	60.42	68.54	68.23	67.47
Education				
Lower than high school	56.89	65.93	64.89	65.86
High school or above	60.24	68.58	65.96	70.46
Participation in training				
Yes	62.69	70.94	67.15	72.12
No	58.04	66.76	65.18	67.76
Foreign languages ability				
Able	63.57	71.02	69.10	72.05
Not able	57.57	66.52	64.59	67.54
Type of area				
Urban	58.11	67.04	64.92	67.79
Rural	61.17	68.80	67.58	70.81
Subdistrict				
Bantur	56.48	65.36	64.00	67.14
Poncokusumo	58.92	67.43	64.66	70.35
Pujon	59.29	67.93	66.24	67.94

After normalization, the individual tourism awareness scores were averaged and subsequently classified into three categories—low, medium, and high—based on the mean and standard deviation. With a mean TAI score of 58.77 and a standard deviation of 19.33, the classification thresholds were set as shown in Table 8.

Table 8. Classification of TAI

Classification	TAI Category
$TAI < 39.44$	Low
$39.44 \leq TAI < 78.11$	Medium
$TAI \geq 78.11$	High

The choice of the mean plus/minus one standard deviation rule was made to ensure meaningful variation across categories. In social and behavioral research, this approach is widely used to classify continuous scores into ordinal levels because most observations typically fall within one standard deviation of the mean, while those above or below are distinguished as higher or lower levels [28], [32]. Using a wider cutoff, such as two standard deviations, would concentrate the majority of cases in the medium category and leave only extreme values in the high and low groups, thereby reducing analytical differentiation. Thus, the mean plus/minus one standard deviation rule provides a balance between statistical rigor and practical interpretability.

This classification scheme provides a systematic approach to differentiating workers' levels of tourism awareness. Scores below 39.44 indicate low awareness, scores between 39.44 and 78.11 reflect moderate awareness, and scores equal to or above 78.11 represent high awareness. By employing mean and standard deviation as the basis for classification, the categorization ensures statistical rigor and comparability across individuals.

Prior to estimating the ordered logistic regression model, multicollinearity among the categorical predictors was assessed using Generalized Variance Inflation Factor (GVIF), which is an extension of the traditional Variance Inflation Factor (VIF) suitable for models containing categorical predictors [49]. The GVIF in Table 9 showed very low values, with the highest GVIF being 1.19 for Area and 1.10 for Subdistrict. Since all GVIF values were substantially below the common threshold of 5, these results suggest that multicollinearity is not a concern in the model [32].

Table 9. Generalized Variance Inflation Factor

Variable	GVIF
Gender	1.02
Age	1.08
Education	1.08
Participation in training	1.04
Foreign languages ability	1.06
Type of area	1.19
Subdistrict	1.10

Having established the classification of the TAI into three ordered categories—low, medium, and high—the next step was to examine the determinants of awareness using an ordered logistic regression model. The model, based on the Scenario 6, included 868 observations. The ordered logistic regression model yielded a McFadden's Pseudo-R^2 of 0.021. Although this value appears relatively low, it is not uncommon in social science research involving ordinal outcomes, particularly when modeling complex human behaviors such as tourism awareness [42], [50]. For instance, studies on residents' attitudes towards immigration and tourist preferences in rural tourism using ordinal or discrete choice models frequently report low Pseudo-R^2 values, typically ranging from 0.01 to 0.08 [51], [52]. This reflects the inherent difficulty of explaining multifaceted behavioral or attitudinal outcomes with limited socio-demographic predictors. Future research may improve the explanatory power by incorporating additional variables such as income, years of experience in the tourism sector, quality and frequency of training programs, and level of exposure to tourist interactions. Nevertheless, the overall model was statistically significant ($\mathrm{LR}\ \chi^2_8 = 29.97, \mathrm{p} < 0.001$), and several key predictors showed statistically significant associations. Thus, the model remains useful for identifying important determinants of tourism awareness, even if a substantial portion of variance remains unexplained. The proportional odds assumption was tested using the Brant test, yielding a chi-square value of 7.796 ($\mathrm{df} = 8, \mathrm{p} = 0.454$). Since the p-value was greater than 0.05, the assumption holds, validating the use of the ordered logistic regression model.

The estimated ordered logistic regression model can be expressed as

$$\log\left(\frac{P(TAI \le j)}{P(TAI > j)}\right) = \theta_j + (0.04\ Gender + 0.35\ Age + 0.47\ Education \\ -0.17\ Training - 0.56\ Foreign + 0.37\ Area \\ -0.08\ Subdistrict_{Poncokusumo} + 0.31\ Subdistrict_{Pujon} \quad (11)$$

for thresholds $j = 1,2$.

Table 10. Ordered Logistic Regression Results for the TAI

Variable	Coefficient	Standard Error	z	p-value	95% Confidence Interval		Odds Ratio
Gender							
Male	-						
Female	0.04	0.15	0.29	0.77	-0.25	0.34	1.04
Age							
15-44	-						
>=45	0.35	0.17	2.03	0.04	0.01	0.69	1.42
Education							
Lower than high school	-						
High school or above	0.47	0.16	2.91	0.00	0.15	0.78	1.60
Participation in training							
Yes							
No	-0.17	0.21	-0.81	0.42	-0.57	0.24	0.85
Foreign languages ability							
Able	-						
Not able	-0.56	0.19	-2.94	0.00	-0.94	-0.19	0.57
Type of area							
Urban	-						
Rural	0.37	0.21	1.79	0.07	-0.03	0.78	1.45
Subdistrict							
Bantur	-						
Poncokusumo	-0.08	0.24	-0.34	0.74	-0.55	0.39	0.92
Pujon	0.31	0.22	1.37	0.17	-0.13	0.75	1.36
Threshold 1	-2.04	0.34			-2.70	-1.37	
Threshold 2	1.53	0.34			0.87	2.19	

Unlike binary or linear regression models, the ordered logit model does not include a single intercept term (β_0). Instead, it estimates two cut points (θ_1 and θ_2) that separate the three ordinal categories on the underlying latent scale. In this model, the estimated cut points are θ_1=-2.04 and θ_2 =1.53. These thresholds define the boundaries between the low, medium, and high categories of tourism awareness.

In line with the second interpretive approach described in the Research Method, an odds ratio indicates the multiplicative change in the odds of being in a higher TAI category versus the combined lower categories, holding other variables constant. Regression coefficients and odds ratios in Table 10 show that several predictors are statistically significant. Workers aged 45 years and above are more likely to be in higher TAI categories ($\beta = 0.35, 95\%\ \text{CI}\ [0.01, 0.69], p = 0.042, \text{OR} = 1.42$). The confidence interval for this coefficient is relatively narrow and lies entirely above zero, indicating a stable positive effect. Education also plays a positive role: those with at least a high school education have significantly higher odds of being in higher categories ($\beta = 0.47, 95\%\ \text{CI}\ [0.15, 0.78], p = 0.004, \text{OR} = 1.60$). The confidence interval is narrow and does not include zero, suggesting a robust positive association. Conversely, lack of foreign language proficiency reduces the likelihood of higher awareness ($\beta = -0.56, 95\%\ \text{CI}\ [-0.94, -0.19], p = 0.003, \text{OR} = 0.57$); the confidence interval is relatively narrow and lies entirely below zero, confirming a consistent negative effect. Rural residence shows a positive but only marginally significant effect ($\beta = 0.37, 95\%\ \text{CI}\ [-0.03, 0.78], p = 0.07, \text{OR} = 1.45$), with a wider confidence interval that includes zero, indicating greater uncertainty around this estimate. Sex, training participation, and subdistrict did not exhibit statistically significant associations.

Although the TAI demonstrates adequate reliability, validity, and robustness across multiple methodological scenarios, several limitations should be noted. The factor structure was derived solely from data collected in three subdistricts of Malang Regency. Although principal component extraction provided an objective, data-driven approach to weighting, the resulting structure remains empirical and context-specific. A high statistical loading does not necessarily imply greater theoretical importance in tourism development. Furthermore, the cultural specificity of the Sapta Pesona framework—developed within the Indonesian policy context—likely influenced the observed clustering of dimensions (particularly Safety–Orderliness–Cleanliness). While this enhances the index's relevance for Indonesia, it also limits its direct generalizability to other cultural or national contexts.

In addition, the study lacks external validation. The TAI has not yet been tested in other tourism destinations or correlated with independent outcome measures such as tourist satisfaction scores or service quality ratings. The cross-sectional design also prevents causal inference and assessment of temporal stability. Future research should conduct pilot testing of the TAI in additional regions with diverse tourism characteristics (e.g., coastal, urban, or heritage destinations) to evaluate the stability and generalizability of the three-factor structure. External validation could be strengthened by examining correlations between TAI scores and objective indicators, such as tourist satisfaction surveys or service quality assessments.

4. CONCLUSION

This study developed and validated the Tourism Awareness Index (TAI) based on the Sapta Pesona framework and analyzed its determinants using ordered logistic regression. Through rigorous reliability and validity testing, factor analysis with principal component extraction, and robustness checks across nine methodological scenarios, the study successfully reduced the original seven dimensions into three coherent and stable factors. The regression analysis further revealed that education, age, and foreign language ability are significant determinants of higher tourism awareness, while rural residence showed a marginally significant positive association. Methodologically, this research advances composite index construction in tourism studies by integrating systematic uncertainty analysis.

While the findings offer valuable insights for workforce development in Malang Regency, they should be interpreted cautiously given the study's geographic scope and cross-sectional design. Local tourism stakeholders may consider prioritizing education and foreign language training programs, as well as leveraging the relatively high awareness levels observed in rural areas. Overall, the TAI provides a transparent and robust tool for measuring individual-level tourism awareness, contributing to methodological development in applied statistics and evidence-based strategies for sustainable tourism.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the support provided by the Center for Research and Community Service (PPPM) of Politeknik Statistika STIS, whose assistance was instrumental in the completion of this study. We also appreciate the institutional resources and guidance that facilitated the research process. This research received no external funding.

5. REFERENCES

[1] UNWTO, "The economic contribution of tourism and the impact of COVID-19," 2021.
[2] BPS-Statistics Indonesia, "Number of Foreign Tourist Visits per Month by Passport," 2024.
[3] OECD, *OECD Tourism Trends and Policies 2022*. Paris: OECD Publishing, 2022.
[4] Q. B. Baloch *et al.*, "Impact of tourism development upon environmental sustainability: a suggested framework for sustainable ecotourism," *Environmental Science and Pollution Research*, pp. 5917–5930, 2023, doi: 10.1007/s11356-022-22496-w.
[5] World Economic Forum, "Travel & Tourism Development Index 2021: Rebuilding for a Sustainable and Resilient Future," Geneva, Switzerland, 2022.
[6] Y. P. Masan *et al.*, "Indeks Pembangunan Kepariwisataan Nasional," Jakarta, 2023.
[7] A. Maulana, M. Akbar, I. Hermawan, A. Wicaksono, and Y. Isnasari, "Indeks Pembangunan Kepariwisataan Nasional (IPKN)," Jakarta, 2024.
[8] M. Nardo, M. Saisana, A. Saltelli, S. Tarantola, A. Hoffman, and E. Giovannini, *Handbook on Constructing Composite Indicators and User Guide*. OECD, 2005.
[9] OECD, *Handbook on Constructing Composite Indicators: Methodology and User Guide*. Ispra, Italy: OECD Publishing, 2008. doi: 10.1007/s12597-023-00703-7.
[10] D. E. K. Boutchich, "Tourism performance evaluation and analysis from composite index and slack based method," *OPSEARCH*, vol. 61, pp. 501–523, 2024.
[11] M. A. Selim, N. A. Abdel-Fattah, and Y. S. Hegazi, "A Composite Index to Measure Smartness and Competitiveness of Heritage Tourism Destination and Historic Building," *Sustainability*, vol. 13, no. 24, pp. 1–24, 2021, doi: 10.3390/su132413135.
[12] K. V Otokiti, K. Gourzis, and T. Kizos, "Touristification dynamics in Greek Islands: a composite index approach," *Journal of Hospitality and Tourism Horizons*, pp. 1–24, 2026, doi: 10.1108/JHTH-08-2025-0114.
[13] K. V Otokiti, "Comparative assessment of vulnerability among small island developing states using a composite index," *Critical Insights in Climate Change*, vol. 2, no. 1, pp. 1–13, 2026, doi: 10.1080/29931495.2026.2646250.
[14] R. Cataldo, M. G. Grassia, and M. Marino, "Composite Indicators for Sustainable Tourism," in *Methodological and Applied Statistics and Demography II*, Bari, Italy, 2024. doi: 10.1007/978-3-031-64350-7_45.
[15] F. J. Blancas and M. Lozano-Oyola, "Sustainable tourism evaluation using a composite indicator with different compensatory levels," *Environ. Impact Assess. Rev.*, vol. 93, 2022, doi: 10.1016/j.eiar.2021.106733.
[16] D. D. Marchi, R. Becarelli, and L. D. Sarli, "Tourism Sustainability Index: Measuring Tourism Sustainability Based on the ETIS Toolkit, by Exploring Tourist Satisfaction via Sentiment Analysis," *Sustainability*, vol. 14, no. 13, pp. 1–18, 2022, doi: 10.3390/su14138049.
[17] A. M. Bindawas, "Promoting Sustainable Tourism Through Employee Skills: Contextualizing Quality Education and the Human Resource Management Perspective (SDG-4)," *Sustainability*, vol. 17, no. 2, 2025, doi: 10.3390/su17020748.
[18] Wahyuningrat and B. T. Harsanto, "Managing human resource in the rural tourism sector: evidence from Indonesia," *Cogent Business & Management*, vol. 12, no. 1, pp. 1–22, 2025, doi: 10.1080/23311975.2025.2493312.
[19] P. Dang, J. Li, and L. Zhang, "Factors Determining Rural Household Participation in Tourism: A Logistic-ISM Model Approach," *Sage Open*, 2026, doi: 10.1177/21582440261421989.
[20] Ministry of Culture and Tourism, "Peraturan Menteri Kebudayaan dan Pariwisata Nomor 4 Tahun 2008 tentang Sadar Wisata," Jakarta, 2008.
[21] R. Azen and C. M. Walker, *Categorical Data Analysis for the Behavioral and Social Sciences*, 2nd ed. New York: Routledge, 2021.
[22] BPS-Statistics of Malang Regency, "Statistik Daerah Kabupaten Malang 2023," Malang, 2023.
[23] J. W. Creswell and J. D. Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, 6th ed. Los Angeles: Sage Publications, 2023.
[24] A. Field, *Discovering Statistics Using IBM SPSS Statistics*, 6th ed. Brighton: SAGE Publications Ltd, 2024.
[25] J. Cohen, P. Cohen, S. G. West, and L. S. Aiken, *Applied Multiple Regression Correlation Analysis for the Behavioral Sciences*, 3rd ed. New York: Routledge Academic, 2002.
[26] R. F. DeVellis and C. T. Thorpe, *Scale Development: Theory and Applications*, 5th ed. Los Angeles: SAGE Publications, Inc, 2021.
[27] L. J. Cronbach, "Coefficient alpha and the internal structure of tests," *Psychometrika*, vol. 16, pp. 297–334, 1951, doi: 10.1007/BF02310555.
[28] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, 6th ed. Upper Saddle River: Pearson Prentice Hall, 2007.

[29] H. F. Kaiser, "An index of factorial simplicity," *Psychometrika*, vol. 39, pp. 31–36, 1974, doi: 10.1007/BF02291575.
[30] G. D. Hutcheson and N. Sofroniou, *The multivariate social scientist: introductory statistics using generalized linear models*. London: SAGE Publications Ltd, 1999.
[31] M. S. Bartlett, "A Note on the Multiplying Factors for Various χ^2 Approximations," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 16, no. 2, pp. 296–298, 1954, doi: 10.1111/j.2517-6161.1954.tb00174.x.
[32] J. F. Hair, W. C. Black, B. J. Babin, and R. E. Anderson, *Multivariate Data Analysis*, 8th ed. Andover: Cengage Learning EMEA, 2019.
[33] A. B. Costello and J. Osborn, "Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most from Your Analysis," *Practical Assessment Research & Evaluation*, vol. 10, pp. 1–9, 2005, doi: 10.7275/jyj1-4868.
[34] M. W. Watkins, "Exploratory Factor Analysis: A Guide to Best Practice," *Journal of Black Psychology*, vol. 44, no. 3, pp. 219–246, 2018, doi: 10.1177/0095798418771807.
[35] B. G. Tabachnick and L. S. Fidell, *Using Multivariate Statistics*, 7th ed. New York: Pearson Education, Inc., 2019.
[36] M. W. Browne, "An Overview of Analytic Rotation in Exploratory Factor Analysis," *Multivariate Behav. Res.*, vol. 36, no. 1, pp. 111–115, 2001, doi: 10.1207/S15327906MBR3601_05.
[37] M. Saisana, A. Saltelli, and S. Tarantola, "Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators," *J. R. Stat. Soc. Ser. A Stat. Soc.*, vol. 168, pp. 307–323, 2005, doi: 10.1111/j.1467-985X.2005.00350.x.
[38] L. Salvati and M. Carlucci, "A composite index of sustainable development at the local scale: Italy as a case study," *Ecol. Indic.*, vol. 43, pp. 162–171, 2014, doi: 10.1016/j.ecolind.2014.02.021.
[39] N. N. Efendi and T. Sirait, "Constructing Out-Migration Potency Index from Province of Indonesia in 2019," *BAREKENG: Journal of Mathematics and Its Applications*, vol. 16, no. 4, pp. 1249–1258, 2022, doi: 10.30598/barekengvol16iss4pp1249-1258.
[40] E. Tate, "Uncertainty Analysis for a Social Vulnerability Index," *Annals of the Association of American Geographers*, vol. 103, no. 3, pp. 526–543, 2013, doi: 10.1080/00045608.2012.700616.
[41] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, 5th ed. New York: Chapman and Hall/CRC, 2020.
[42] D. W. Hosmer Jr., S. Lemeshow, and R. X. Sturdivant, *Applied Logistic Regression*. Hoboken, New Jersey: John Wiley & Sons, Inc., 2013.
[43] D. G. Kleinbaum and M. Klein, *Logistic Regression: A Self-Learning Text*. New York: Springer, 2010.
[44] P. McCullagh, "Regression Models for Ordinal Data," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 42, no. 2, pp. 109–142, 1980, doi: 10.1111/j.2517-6161.1980.tb01109.x.
[45] R. Brant, "Assessing Proportionality in the Proportional Odds Model for Ordinal Logistic Regression," *Biometrics*, vol. 46, no. 4, pp. 1171–1178, 1990, doi: 10.2307/2532457.
[46] B. Peterson and F. E. Harrell Jr., "Partial Proportional Odds Models for Ordinal Response Variables," *J. R. Stat. Soc. Ser. C Appl. Stat.*, vol. 39, no. 2, pp. 205–217, 1990, doi: 10.2307/2347760.
[47] J. W. Hardin and J. M. Hilbe, *Generalized Linear Models and Extensions*, 4th ed. College Station: Stata Press, 2018.
[48] J. W. Grice, "Computing and evaluating factor scores," *Psychol. Methods*, vol. 6, no. 4, pp. 430–450, 2001, doi: 10.1037/1082-989X.6.4.430.
[49] J. Fox, *Applied regression analysis and generalized linear models*, 3rd ed. Thousand Oaks, CA: Sage, 2016.
[50] J. S. Long and J. Freese, *Regression Models for Categorical Dependent Variables Using Stata*, 3rd ed. College Station: Stata Press, 2014.
[51] A. Ivlevs and I. Smith, "Do international tourist arrivals change residents' attitudes towards immigration? A longitudinal study of 28 European countries," *Tourism Economics*, vol. 30, no. 1, pp. 104–131, 2024, doi: 10.1177/13548166221137950.
[52] W. An and S. Alarcón, "Rural tourism preferences in Spain: Best-worst choices," *Ann. Tour. Res.*, vol. 89, pp. 1–13, 2021, doi: 10.1016/j.annals.2021.103210.