Annals of Tropical Research 47 (2):201-217(2025)
https://doi.org/10.32945/atr47214.2025

ANNALS OF TROPICAL RESEARCH

Original Article

Factor structure validation of the nurses' behavior toward confirmed and suspected HIV/AIDS patients (NB-CSHAP) scale

Norberto E. Milla Jr.[1*] and Janet Alexis A. De los Santos[2]

ABSTRACT

Behavioral responses of nurses toward people living with HIV are critical determinants of care quality, patient trust, and treatment outcomes. Stigma-driven practices continue to undermine ethical and inclusive healthcare delivery in many clinical settings. However, limitations in validated behavioral measurement tools hinder the systematic evaluation of inclusive and discriminatory nursing behaviors. This study validates the factor structure of the Nurses' Behavior Toward Confirmed and Suspected HIV/AIDS Patients (NB-CSHAP) scale, a novel instrument developed to assess the spectrum of inclusive and discriminatory service behaviors in stigma-sensitive healthcare settings. We evaluated the scale using responses from a sample of 400 nurses with direct experience in caring for suspected or confirmed HIV/AIDS cases. The psychometric properties were rigorously tested using Confirmatory Factor Analysis (CFA) with five-fold cross-validation. Model fit statistics (CFI=0.905, RMSEA=0.071, SRMR=0.070) confirmed that the hypothesized four-factor model provides a good fit to the data, and standardized factor loadings supported item convergence. Internal consistency was strong for the overall scale, with composite reliability (CR) values ranging from 0.459 to 0.827. The Discriminatory factor showed the highest internal consistency and convergent validity. Moreover, HTMT ratios confirmed discriminant validity across all latent constructs. Factor covariances revealed theoretically consistent relationships, particularly the expected inverse association between the Service-Oriented and Discriminatory behaviors. While most factors performed well, the Perceptiveness and Openhandedness factors demonstrated weaker reliability and convergence, suggesting the need for targeted item refinement in future research. The NB-CSHAP scale is a contextually grounded and psychometrically sound tool for measuring behavioral diversity in sensitive healthcare settings, making it highly valuable for guiding stigma-reduction interventions and research in HIV care.

Keywords: Confirmatory factor analysis, HIV, latent construct, psychometric validation, scale development

Received: 5 October 2025
Revised: 4 December 2025
Accepted: 5 December 2025
Published: 12 December 2025

©The authors. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 (https://creativecommons.org/licenses/by-nc-nd/4.0/)

[1]Department of Statistics, Visayas State University, Visca, Baybay City, Leyte, 6521-A, Philippines
[2]Faculty of Nursing, Visayas State University, Visca, Baybay City, Leyte, 6521-A, Philippines

*Corresponding Author. Address: Department of Statistics, Visayas State University, Visca, Baybay City, Leyte, 6521-A, Philippines; Email: bertmilla@vsu.edu.ph

INTRODUCTION

Stigma in HIV/AIDS care remains a persistent barrier to equitable health outcomes, particularly in tropical regions where healthcare systems face resource constraints and sociocultural complexities (Gurung et al., 2025; Relf et al., 2021; Zhang et al., 2024). Nurses, as frontline providers, play a pivotal role in shaping the care experiences of people living with HIV/AIDS (Ngcobo et al., 2024). However, discriminatory attitudes and behaviors in healthcare settings can significantly undermine trust, reduce service uptake, and perpetuate health disparities (Koseoglu Ornek et al., 2020; WHO, 2023).

Despite global efforts to reduce HIV-related stigma, few validated instruments exist to assess nurse behavior in clinical settings, especially in low- and middle-income tropical countries. Many existing tools lack cultural sensitivity, multidimensional design, and psychometric robustness, limiting their utility in organizational assessments, stigma-responsive training, and SDG-aligned reporting (Gurung et al., 2025; Luz et al., 2020).

To address this gap, the Nurse Behavior toward Confirmed and Suspected HIV/AIDS Patients (NB-CSHAP) scale was developed using a sequential exploratory mixed-method design, integrating qualitative insights from PLHIV and expert input from nursing professionals (De Los Santos et al., 2022). The NB-CSHAP scale assesses nurses' behavioral tendencies, both affirming and discriminatory, toward individuals living with or suspected of having HIV/AIDS. Designed for use in low-resource healthcare settings, the scale focuses on observable behaviors rather than attitudes or knowledge alone, aligning with best practices in culturally grounded scale development and participatory psychometrics (DeVellis & Thorpe, 2022).

The scale comprises 16 items grouped into four latent constructs: (1) *Service-oriented*, measuring respectful and patient-centered behaviors; (2) *Discriminatory*, capturing avoidance and exclusionary tendencies; (3) *Openhandedness*, reflecting generosity and nonverbal warmth; and (4) *Perceptiveness*, assessing empathy and attentiveness to patient needs. With an internal consistency of 0.73, the NB-CSHAP scale demonstrates acceptable psychometric properties. However, further validation is essential before its routine deployment in clinical settings (De Los Santos et al., 2022).

A critical phase in scale development is the assessment of factorial validity, which refers to the extent to which a proposed factor structure accurately represents the underlying constructs. Confirmatory Factor Analysis (CFA), a theory-driven statistical technique, is commonly used to evaluate model fit and latent structure (Hair et al., 2022; Kline, 2023). Establishing factorial validity is vital not only for confirming the internal structure of the NB-CSHAP scale but also for ensuring its generalizability across diverse populations and settings. Moreover, factorial validity lays the foundation for subsequent psychometric evaluations, such as measurement invariance, predictive validity, and responsiveness to interventions, which are key to applying the scale in organizational reviews, stigma-awareness training programs, and SDG-aligned documentation (Elsman et al., 2024; Gurung et al., 2025).

This study presents the CFA and psychometric evaluation of the NB-CSHAP scale. Specifically, it examines the factorial structure, model fit indices, reliability

coefficients, and validity evidence of the scale. By validating this tool, the study contributes to evidence-based advocacy and capacity-building in healthcare systems, in line with the Global Health Sector Strategies on HIV (2022–2030) (WHO, 2022), and the 2030 Agenda for Sustainable Development (United Nations, 2015).

MATERIALS AND METHODS

Data

The data set used in this paper came from the study of De Los Santos et al. (2022) in which 400 nurses working in various health facilities participated in the survey. They were selected based on these criteria: (a) registered nurses who are in service for at least 6 months; and (b) have experience in caring for either a suspected or confirmed case of HIV/AIDS. The participants were asked to complete the first version of the NB-CSHAP instrument, which initially consisted of 16 items.

Statistical Analysis

Model specification

The hypothesized four-factor model of the NB-CSHAP scale consisted of 16 items grouped into the following categories: *Service-Oriented* (7 items), *Discriminatory* (4 items), *Openhandedness* (3 items), and *Perceptiveness* (2 items). Each item was specified to load on its respective latent construct, with no cross-loadings. CFA was conducted with five-fold cross-validation to assess model stability. All results presented in the succeeding sections are based on 5-fold cross-validation. The hypothesized four-factor model is shown below.

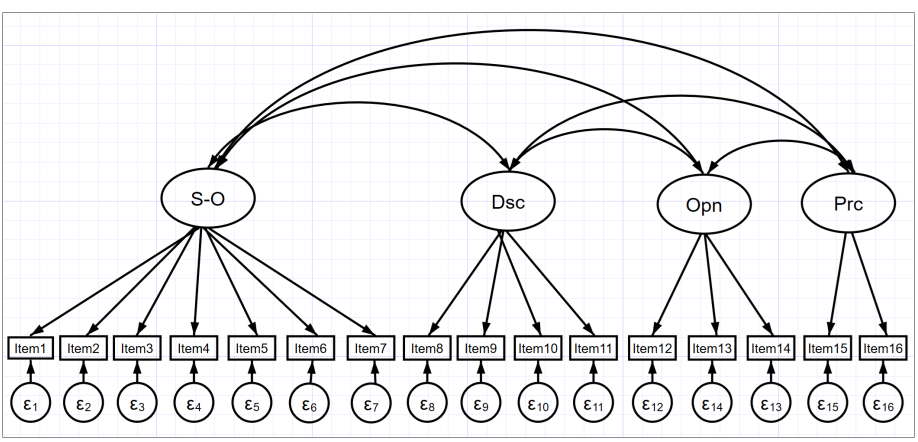

Figure 1. The Hypothesized Four-Factor Model of the NB-CSHAP

Factorial validity refers to the extent to which a set of observed variables accurately reflects the underlying theoretical constructs they are intended to measure. In scale development, establishing factorial validity is essential to

ensure that each item loads meaningfully onto its designated factor and that the overall structure of the instrument aligns with theoretical expectations (DeVellis & Thorpe, 2022).

CFA is the primary statistical method used to assess factorial validity. Unlike Exploratory Factor Analysis (EFA), which seeks to uncover latent structures without prior assumptions, CFA tests a predefined factor model based on theory or prior empirical evidence (Kline, 2023). It evaluates how well the hypothesized model fits the observed data. In this particular study, we are testing whether the four-factor model derived from exploratory factor analysis can be supported by the data. Various statistical measures are used to assess the factorial validity of the NB-CSHAP scale.

Validation Metrics and Standard Output of CFA

The quality of fit of the four-factor model is evaluated using the following metrics:

1.Model fit indices. These include the Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square Residual (SRMR). Values of CFI and TLI above 0.95 indicate excellent fit, and values above 0.90 are considered acceptable in early-stage validation. Meanwhile, RMSEA, which estimates how well a hypothesized model would fit the population covariance matrix, not just the sample data, is interpreted as a *Close fit* (<0.05), *Reasonable fit* (0.05–0.08), *Mediocre fit* (0.081–0.10), and *Poor fit* (>0.10) (Hair et al., 2022; S & Mohanasundaram, 2024). SRMR reflects the average deviation between the actual and estimated correlations produced by the model. It is an absolute fit index that is scale-independent, making it useful for comparing models across different datasets or constructs. According to Kline (2023), its value can be interpreted as an *Acceptable fit* (≤0.08) and a Good fit (≤0.05). Together, RMSEA and SRMR a provide robust picture of how well the model reproduces the observed data.

2. Standardized factor loadings. These represent the strength and direction of the relationship between an observed variable (item) and its underlying latent construct (factor) in CFA. They indicate how well each item reflects the factor it is intended to measure (Kline, 2023; S & Mohanasundaram, 2024). High standardized loadings indicate that the item makes a meaningful contribution to the latent construct and supports convergent validity, which is the idea that items within a factor measure the same concept (Hair et al., 2022). Standardized loadings above 0.7 indicate strong item-factor correlation, standardized loadings between 0.50 and 0.69 indicate moderate item-factor correlation, and standardized loadings less than 0.50 represent weak correlation between an item and the corresponding factor (S & Mohanasundaram, 2024).

3. Factor variances and covariances. Factor variances represent the amount of variability in each latent construct within a CFA model. They indicate the degree to which individuals differ in their scores on a given factor. A higher

factor variance suggests that the construct is meaningfully differentiated across respondents, while a very low variance may imply limited variability or redundancy in measurement (Kline, 2023; S & Mohanasundaram, 2024). Factor covariances reflect the degree to which two latent constructs vary together. In CFA, they help assess discriminant validity, whether constructs are distinct yet appropriately related. A positive covariance indicates that as one factor increases, the other tends to increase as well; a negative covariance suggests an inverse relationship (Kline, 2023). Covariances are essential for understanding the structural coherence of the model and for evaluating whether the latent constructs behave as theorized (Hair et al., 2022; S & Mohanasundaram, 2024).

4. Validity and Reliability Measures. The Average Variance Extracted (AVE) measures the amount of variance captured by a latent construct in relation to the variance due to measurement error. It is used to assess convergent validity, which refers to how well items are intended to measure the same construct actually do so (Cheung et al., 2024) In other words, convergent validity indicates how strongly the NB-CSHAP scale correlates with other scales that measure the same or similar construct. Simply, it is a measure of "similarity" or "agreement" with other similar scales. AVE of 0.50 and above indicates that the construct explains at least half of the variance in its indicators, supporting convergent validity, and AVE < 0.50 suggests that error variance dominates, and the construct may not be well represented by its items. The Heterotrait-Monotrait Ratio (HTMT) is a modern and robust method for assessing discriminant validity, which evaluates whether constructs that are supposed to be distinct are indeed empirically different. Discriminant validity indicates how distinct or different the NB-CSHAP scale is from other scales that measure different constructs. HTMT < 0.85 suggests strong evidence of discriminant validity; HTMT < 0.90 can be considered acceptable discriminant validity, and HTMT ≥ 0.90 indicates potential overlap between constructs; discriminant validity may be compromised (Roemer et al., 2021). Composite reliability (CR) is a measure of the internal consistency of a latent construct, indicating how well a set of observed variables (items) reliably measures the same underlying factor. While Cronbach's alpha has traditionally been used for this purpose, omega (ω) is now widely recommended as a more accurate and flexible alternative, especially in CFA and structural equation modeling contexts (Cheung et al., 2024). Omega accounts for unequal factor loadings, whereas alpha assumes all items contribute equally to the construct. It is less biased when items vary in strength or when the scale includes fewer items. Omega is derived from the CFA model, utilizing standardized loadings and error variances, making it an ideal choice for latent variable modeling (Cheung et al., 2024). Values of omega at least 0.70 indicate acceptable internal consistency, at least 0.80 implies good reliability, at least 0.90 means an excellent reliability, and values below 0.70 may indicate poor consistency or a need for item revision (Cheung et al., 2024; Hair et al., 2022; S & Mohanasundaram, 2024).

K-fold Cross Validation

K-fold cross-validation is a resampling technique used to assess the generalizability and stability of statistical models, particularly in predictive modeling and psychometric validation. It works by partitioning the dataset into *k* equally sized subsets (or "folds"). In k-fold cross-validation, the model is trained on all but one fold and tested on the excluded fold. This cycle is repeated k times, ensuring that each fold functions as the test set exactly once. The results are then averaged to produce a more robust estimate of model performance (James et al., 2021; Teodorescu & Obreja Brașoveanu, 2025).

In psychometric research, k-fold validation helps evaluate the stability of factor structures, reliability indices, and fit statistics across different subsamples. This is especially important when validating instruments like the NB-CSHAP scale, where consistency across folds strengthens claims of factorial validity, internal consistency, and generalizability (James et al., 2021; Teodorescu & Obreja Brașoveanu, 2025). Compared to a single train-test split, k-fold validation reduces the chance of overfitting and provides a more accurate estimate of how the model will perform on new data (Aguilar-Ruiz, 2025). Typical choices of *k* are 5 or 10. Larger *k* values reduce bias but may increase variance; smaller *k* values reduce variance but may increase bias (James et al., 2021). In this study, a 5-fold cross-validation was implemented, and all fit statistics and measures of validity and reliability were averaged across folds.

Although the initial Exploratory Factor Analysis (EFA) was performed on the full dataset to establish the preliminary factor structure, Confirmatory Factor Analysis (CFA) combined with a five-fold cross-validation was subsequently employed on the same dataset. This approach allows us to formally test the structural validity of the EFA-derived model and rigorously assess its generalizability and stability across different subsets of the data, providing a more robust validation than a single-sample CFA. By doing this, possible bias and overfitting from using the same data set are minimized.

Software Used

Confirmatory factor analysis, including the 5-fold cross-validation, was implemented in R v4.5.1, RStudio 2025.09.0+387, and JASP v0.95.2. (JASP Team, 2025; Posit team, 2025; R Core Team, 2025).

RESULTS

Model Fit Indices

The estimated four-factor structure of the NB-CSHAP scale is presented in Figure 2. To assess the fit of this factor structure to the data, the standard fit indices described in the Methods section are reported in Table 1. The reported values are based on the 5-fold cross-validation CFA. The CFI value of 0.905 indicated an acceptable fit, but a TLI of 0.884 fell slightly below the conventional cutoff of 0.90 set by Hair et al. (2022), Kline (2023), and S & Mohanasundaram (2024).

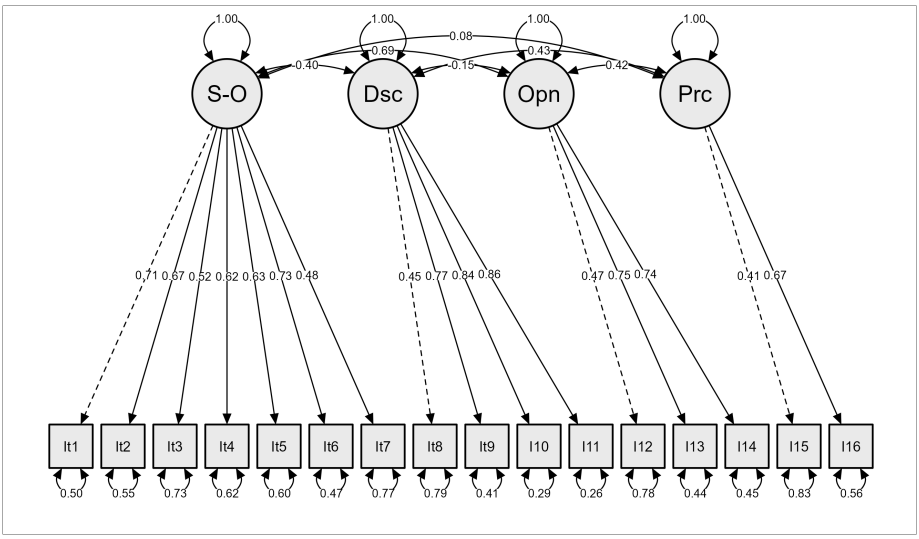

Figure 2. Estimated Four-factor Structure of the NB-CSHAP Scale

Table 1. Fit Indices for the Confirmatory Factor Analysis of the NB-CSHAP Scale

Fit Indices	Value	Interpretation
Comparative Fit Index (CFI)	0.905	Acceptable fit
Tucker-Lewis Index (TLI)	0.884	Marginal fit
Root mean square error of approximation (RMSEA)	0.071	Reasonable fit
Standardized root mean square residual (SRMR)	0.070	Acceptable fit

The other fit measures are RMSEA = 0.071, and SRMR = 0.070. These values both indicate a reasonable and acceptable fit, respectively, according to the guidelines of Hair et al. (2022), Kline (2023), and S & Mohanasundaram (2024).

Standardized Factor Loadings

The standardized loadings are shown in Table 2. In CFA, standardized factor loadings reflect the strength of the relationship between each observed item and its underlying latent construct. These loadings are analogous to correlation coefficients and range from –1 to +1. Higher values indicate that an item is a strong indicator of the latent factor it is intended to measure. In other words, those items are statistically and conceptually aligned with the factor they're intended to measure.

Standardized factor loadings for the *Service-oriented* latent factor range from 0.482 to 0.726. Item 1 (0.711) and Item 6 (0.726) have loadings above the commonly accepted threshold of 0.70, which suggests that these items share a high proportion of variance with the *Service-Oriented* factor. On the other hand, the standardized loading of Item 7 (0.482) falls below the ideal threshold. The other items (2, 3, 4, and 5) exhibit moderate correlation with the latent factor *Service-Oriented*.

For the *Discriminatory* factor, the standardized loading of Item 8 (0.450) falls below the commonly accepted threshold of 0.50. The rest of the items in the *Discriminatory* factor have strong (acceptable) loadings.

Table 2. Standardized Factor Loadings for Confirmatory Factor Analysis of the NB-CSHAP Scale

Factor	Item	Standardized Loadings	Interpretation
Service-oriented	Item1	0.711	Strong
	Item2	0.677	Moderate
	Item3	0.523	Moderate
	Item4	0.614	Moderate
	Item5	0.630	Moderate
	Item6	0.726	Strong
	Item7	0.482	Weak
Discriminatory	Item8	0.450	Weak
	Item9	0.766	Strong
	Item10	0.838	Strong
	Item11	0.857	Strong
Openhandedness	Item12	0.476	Weak
	Item13	0.745	Strong
	Item14	0.742	Strong
Perceptiveness	Item15	0.412	Weak
	Item16	0.669	Moderate

The standardized loadings of Item 13 (0.745) and Item 14 (0.742) in the *Openhandedness* factor both exceed the commonly accepted threshold of 0.70, but the standardized loading of Item 12 (0.476) falls below the 0.50 threshold.

Finally, the factor loading of Item 15 (0.412) falls below the conventional threshold of 0.50, and Item 16 (0.669) has a loading within the moderate range (0.50–0.69).

Factor Variances and Covariances

In CFA, factor variances reflect the degree of dispersion in latent trait scores across respondents. Higher values indicate greater individual differences in the construct. As shown in Table 3, the *Discriminatory* factor shows the highest variance (0.406), indicating substantial individual differences in discriminatory attitudes or behaviors among nurse respondents. Meanwhile, the other factors show relatively lower and similar variances, ranging from 0.262 to 0.268, suggesting more consistent responses across individuals, possibly due to shared professional norms or training in healthcare settings.

Table 3. Estimated Factor Variances for the NB-CSHAP Scale (Unstandardized Solution)

Factor	Variance
Service-Oriented	0.265
Discriminatory	0.406
Openhandedness	0.268
Perceptiveness	0.262

As shown in Table 4, many latent constructs of the NB-CSHAP scale reveal positive covariances between several latent constructs, including *Service-Oriented* and *Openhandedness* (0.183), and *Openhandedness* and *Perceptiveness* (0.106), while, a few constructs exhibit negative covariances, most notably between *Service-Oriented* and *Discriminatory* behaviors (−0.115), and to a lesser extent between *Discriminatory* and *Openhandedness* (−0.032).

Table 4. Estimated Factor Covariances for the NB-CSHAP Scale (Unstandardized Solution)

Factors	Covariance
Service-Oriented ↔ Discriminatory	-0.115
Service-Oriented ↔ Openhandedness	0.183
Service-Oriented ↔ Perceptiveness	0.051
Discriminatory ↔ Openhandedness	-0.032
Discriminatory ↔ Perceptiveness	0.115
Openhandedness ↔ Perceptiveness	0.106

Validity and Reliability Measures

Tables 5 and 6 present estimated measures of reliability and validity of the NB-CSHAP scale. The average variance extracted (AVE) is used both as a measure of convergent validity and construct reliability, while the composite reliability, based on the omega coefficient, is used to measure the internal consistency of the items in the scale. As shown in Table 5, the *Discriminatory* factor demonstrates strong psychometric properties, with AVE = 0.556 and CR = 0.827. The *Service-Oriented* factor shows acceptable reliability (CR = 0.819), though its AVE (0.396) falls below the 0.50 benchmark, indicating weak convergence. *Openhandedness* yields borderline reliability (CR = 0.698) and suboptimal AVE (0.444), which suggests the need for item refinement. The *Perceptiveness* factor exhibits the weakest performance, with AVE = 0.309 and CR = 0.459, suggesting poor item coherence and insufficient construct representation.

Table 5. Convergent validity and composite reliability of the NB-CSHAP scale

Factors	Average Variance Extracted (AVE)		Composite Reliability (CR)	
	Value	Interpretation	Value	Interpretation
Service-Oriented	0.396	Weak	0.819	Acceptable
Discriminatory	0.556	Acceptable	0.827	Acceptable
Openhandedness	0.444	Weak	0.698	(Borderline) acceptable
Perceptiveness	0.309	Weak	0.459	Low

Discriminant validity ensures that measures of different constructs are unique and do not significantly overlap. A value less than 0.85 indicates discriminant validity. In this study, the Heterotrait-Monotrait Ratio of Correlations (HTMT) was computed as a measure of discriminant validity. The HTMT values between all latent factors are presented in Table 6. All HTMT values are below the threshold, suggesting the NB-CSAHP scale achieves discriminant validity. The highest HTMT value was observed between *Service-Oriented* and *Openhandedness* (0.656), and the lowest HTMT value was found between *Discriminatory* and *Openhandedness* (0.156).

Table 6. Heterotrait-Monotrait Ratio of Correlations (HTMT) of the factor structure of the NB-CSHAP scale

Factors	HTMT	Interpretation
Service-Oriented ↔ Discriminatory	0.326	Discriminant validity is established
Service-Oriented ↔ Openhandedness	0.656	Discriminant validity is established
Service-Oriented ↔ Perceptiveness	0.219	Discriminant validity is established
Discriminatory ↔ Openhandedness	0.156	Discriminant validity is established
Discriminatory ↔ Perceptiveness	0.525	Discriminant validity is established
Openhandedness ↔ Perceptiveness	0.474	Discriminant validity is established

DISCUSSION

The model fit indices obtained from the CFA of the NB-CSHAP scale suggest that the four-factor structure of the scale reproduces the observed data reasonably well. The acceptable CFI value suggests that the model fits the data reasonably well, though not perfectly. It meets the minimum threshold for acceptability, but further refinement could improve the scale. The RMSEA and SRMR values further support these findings, as both are within the acceptable range. Although the TLI falls slightly below the conventional cutoff of 0.90, indicating a marginal fit, it suggests that the scale may benefit from refinement or simplification (Hair et al., 2022; Kline, 2023; S & Mohanasundaram, 2024). Given the strong performance of the other indices (CFI, RMSEA, SRMR), minor but theoretically sound adjustments, such as rephrasing or rewording a few items or the addition of items to some of the constructs, can be made to push the TLI above 0.90. In a general sense, the CFA provides sufficient evidence to support the validity and structure of the NB-CSHAP Scale.

In CFA, standardized factor loadings reflect the strength of the relationship between each observed item and its underlying latent construct. These loadings are analogous to correlation coefficients and range from –1 to +1. Higher values indicate that an item is a strong indicator of the latent factor it is intended to measure. In other words, those items are statistically and conceptually aligned with the factor they're intended to measure.

Items with loadings above the commonly accepted threshold of 0.70 share a high proportion of variance with their corresponding latent construct. In other words, they are strong indicators of the construct, and they contribute meaningfully to the internal consistency and conceptual clarity of the scale. These high item loadings support convergent validity, internal consistency, and suggest

psychometric strength of the scale. Examples of these items are Items 1 and 6 of the *Service-Oriented* construct, Items 9, 10, and 11 of the *Discriminatory* factor, and Items 13 and 14 of the *Openhandedness* latent construct. Meanwhile, items with standardized loadings below 0.5 have weaker alignment with the latent construct (Items 7, 8, 12, and 15). These items may not be capturing the intended behavior effectively, which could be attributed to ambiguous or culturally sensitive wording, overlap with other constructs, or low variability in responses. These can be rephrased or reworded to improve their correlation with the latent construct (DeVellis & Thorpe, 2022; Gurung et al., 2025). The other items on the scale, such as Items 2, 3, 4, and 5 on the latent factor Service-Oriented, and Item 16 on Perceptiveness factor, which exhibit a moderate loading can also be refined for retention.

Latent construct with high factor variance means that individual responses to the indicator items of this construct are substantially different. High factor variance suggests the construct's sensitivity and relevance in capturing behavioral diversity, especially in stigma-sensitive contexts. The Discriminatory factor of the NB-CSHAP scale showed the highest variance among the four factors, indicating that nurse respondents show largely varying discriminatory behavior. As outlined by DeVellis & Thorpe (2022) and Kline (2023), the observed factor variances reflect meaningful individual differences in latent traits, supporting the NB-CSHAP scale's sensitivity to behavioral diversity in stigma-sensitive healthcare contexts. This is further supported by the study of Pearl et al. (2024), which highlights that stigma-sensitive instruments benefit from multidimensional modeling and context-specific validation. The NB-CSHAP scale similarly reflects behavioral diversity in healthcare interactions, supported by its factor-level dispersion and discriminant structure.

Meanwhile, the other factors Service-Oriented, Openhandedness, and Perceptiveness show relatively lower variance suggesting more consistent responses across individuals, possibly due to shared professional norms or training in healthcare settings. A low factor variance means that the true scores of individuals on the NB-CSHAP construct are clustered closely together. It suggests that there is not much variation in the underlying trait being measured within the sample. Furthermore, the factor with low variance is doing a poor job of discriminating among individuals; that is, it cannot effectively tell one person apart from another on the trait. These factor variances align with recent validation studies of health-related instruments. For example, Wang et al., (2025) reported explained variances between 12.4% and 26.7% across factors in the Health Promotion Literacy Scale, interpreting these as evidence of construct differentiation without excessive heterogeneity. Similarly, Loureiro et al. (2025) found moderate latent dispersion across the emotional, psychological, and social well-being domains in the Mental Health Continuum-Short Form, supporting its multidimensional structure. Baharum et al. (2023) also emphasized balanced latent variability in their second-order CFA of the Newly Employed Nurses Adaptation Questionnaire (NENA-q) model, noting that moderate factor-level dispersion was appropriate for adaptation-related traits in newly employed nurses. Taken together, these findings reinforce the interpretability and structural coherence of the NB-CSHAP scale, particularly in stigma-sensitive healthcare contexts where behavioral diversity must be captured without compromising psychometric stability.

Factor covariances reflect shared variance between latent traits and should be interpreted in light of theoretical expectations. In addition, Kline (2023) notes that moderate covariances support discriminant validity while acknowledging conceptual proximity. This interpretation is reinforced by Wang et al. (2025), who validated the Health Promotion Literacy Scale and reported moderate inter-factor covariances between domains such as Social Interaction and Personal Growth, concluding that these relationships reflect coherent but multidimensional behavioral patterns. In the context of NB-CSHAP, the observed covariances confirm the scale's ability to capture nuanced service behaviors in stigma-sensitive healthcare settings, where traits such as generosity and perceptiveness may reinforce, but not duplicate, compassionate care.

Many of the factors of the NB-CSHAP scale have positive covariances. A positive covariance means that as the values of one factor increase, the values of the other factor also increase. The largest positive covariance is 0.183 for Service-Oriented and Openhandedness. When an individual scores high on the latent trait of Service-Oriented, they tend to also score high on the latent trait of Openhandedness. Conversely, those who score low on Service-Oriented tend to score low on Openhandedness.

The negative covariance, say between Service-Oriented and Discriminatory constructs, indicates inverse relationships. This means that as compassionate, service-oriented behaviors increase, discriminatory tendencies tend to decrease, which is an interpretation consistent with theoretical expectations of behavioral polarity in stigma-sensitive healthcare contexts. DeVellis and Thorpe (2022) emphasize that negative covariances between latent variables reflect theoretical opposition and support discriminant validity, particularly when constructs are expected to diverge in practice. Kline (2023) similarly notes that such covariances indicate not only construct distinctiveness but also behavioral tension. This interpretation is further supported by Pearl et al. (2024), who reported negative covariances between *Stigma Resistance* and *Self-Devaluation* in the Internalized Health-Related Stigma (I-HEARTS) scale, concluding that these patterns reflect meaningful divergence in internalized stigma responses. In the NB-CSHAP framework, the presence of negative covariances reinforces the scale's ability to differentiate inclusive service behaviors from exclusionary attitudes, affirming its structural validity and relevance in equity-focused healthcare research.

In terms of validity and reliability, the NB-CSHAP scale shows generally acceptable validity and reliability. For instance, the AVE and CR values of the *Discriminatory* factor exceed the recommended thresholds for convergent validity and internal consistency. This implies that the *Discriminatory* factor demonstrates robust internal consistency and convergent validity, affirming its structural integrity and relevance in stigma-sensitive contexts. The composite reliability of the *Service-Oriented* factor suggests acceptable reliability, though its average variance explained falls below the 0.50 benchmark, indicating weak convergence. This means that, while the construct is reliable, it shows limited convergence, which may reflect item-level dispersion or conceptual breadth. These results align with findings from Wang et al. (2025), who reported similar AVE–CR discrepancies in their validation of the Health Promotion Literacy Scale, interpreting them as indicators of behavioral nuance rather than measurement failure. *Openhandedness* yields borderline reliability and suboptimal AVE, suggesting the need for refinement of items. The low value of AVE and CR for the *Perceptiveness*

factor suggests poor item coherence and insufficient construct representation. This could point to potential conceptual ambiguity or insufficient item saturation. Adding more items to the *Perceptiveness* construct may increase its reliability and validity. As Rogers (2024) pointed out, reporting AVE and CR together provides a more nuanced view of construct quality, allowing researchers to distinguish between internal consistency and convergent strength. These findings suggest that while the NB-CSHAP scale addresses both supportive and harmful behaviors, further refinement, particularly of the *Perceptiveness* items, is needed to improve psychometric robustness.

The discriminant validity of the NB-CSHAP scale, as reflected by the HTMT values, is below the threshold of 0.85, suggesting satisfactory discriminant validity. In other words, the constructs are distinct or non-redundant. The highest HTMT value was observed between *Service-Oriented* and *Openhandedness*, which may reflect behavioral synergy in compassionate care, an interpretation supported by Wang et al. (2025), who reported similar HTMT values between social interaction and personal growth in their health promotion literacy scale. The lowest HTMT between *Discriminatory* and *Open-handedness* affirms a strong empirical separation between inclusive and exclusionary behaviors, which reinforces the theoretical and empirical distinction between inclusive and exclusionary behaviors. This result aligns with Pearl et al. (2024), who found negative or low HTMT values between stigma resistance and self-devaluation in the I-HEARTS scale and interpreted them as indicators of behavioral polarity in stigma-sensitive contexts. Overall, the HTMT values confirm the structural integrity of the NB-CSHAP scale and its ability to differentiate nuanced service behaviors in healthcare settings.

A primary limitation of the study relates to the psychometric validation of the NB-CSHAP scale. Ideally, the Exploratory Factor Analysis (EFA) and CFA phases should be conducted on two independent, temporally distinct samples to rigorously validate the factor structure of the scale and ensure its generalizability (Brown, 2015). Due to practical constraints on data collection and resource availability, this study employed the same dataset (N=400) that was used in the EFA by De Los Santos et al. (2022). However, a 5-fold cross-validation procedure was implemented to evaluate the stability and generalizability of the factor structure.

CONCLUSION

The findings of this study provide strong support for the four-factor structure of the NB-CSHAP scale, establishing its utility as a psychometrically sound instrument for assessing nurses' behaviors toward confirmed and suspected HIV/AIDS patients. The overall model demonstrates acceptable fit, with theoretically coherent factor covariances and satisfactory discriminant validity, confirming the scale's capacity to distinguish between inclusive and discriminatory service behaviors. The Discriminatory construct shows the most robust performance, while Service-Oriented and Openhandedness factors reflects acceptable reliability with weaker convergence. Perceptiveness emerges as the least stable domain, indicating the need for item refinement and possible expansion to better capture empathy-related behaviors.

Taken together, these results highlight the NB-CSHAP as a contextually grounded tool that extends existing measurement approaches by shifting focus from attitudes to observable clinical behaviors. Its multidimensional framework offers a means to evaluate stigma-sensitive practice, monitor institutional performance, and inform targeted interventions in healthcare systems where HIV stigma remains a persistent barrier to care. Although further refinement is warranted, particularly in domains with weaker reliability, the scale provides a solid foundation for future applications in research, training, and policy development aimed at fostering compassionate and equitable HIV services.

Acknowledgment

The authors wish to thank Visayas State University for the facilities and time provided for the authors in the development of this manuscript.

Author Contributions

NEMJr conceptualized the study. JAADLS gathered the data used for the analysis. Both authors wrote and approved the contents of this manuscript.

Funding Source

This paper was prepared without any financial support.

Availability of Data and Materials

The data and the R script used in the analysis can be shared upon request.

Ethical Considerations

This article made use of secondary data; hence, no human subjects or animals were involved in writing the paper.

Competing Interest

The authors declare no competing interests.

Use of Generative AI

Generative AI was utilized exclusively for the enhancement of grammar and syntax. All intellectual content, analysis and interpretations were developed by the authors.

References

Aguilar-Ruiz, J. S. (2025). *Irredundant k-Fold Cross-Validation* (Version 2). arXiv. https://doi.org/10.48550/arXiv.2507.20048

Baharum, H., Ismail, A., Awang, Z., McKenna, L., Ibrahim, R., Mohamed, Z., & Hassan, N. H. (2023). Validating an instrument for measuring newly graduated nurses'

adaptation. *International Journal of Environmental Research and Public Health, 20*(4), 2860. https://doi.org/10.3390/ijerph20042860

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health, 6*, 149. https://doi.org/10.3389/fpubh.2018.00149

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). The Guilford Press. https://psycnet.apa.org/record/2015-10560-000

Cheung, G. W., Cooper-Thomas, H. D., Lau, R. S., & Wang, L. C. (2024). Reporting reliability, convergent and discriminant validity with structural equation modeling: A review and best-practice recommendations. *Asia Pacific Journal of Management*, *41*(2), 745–783. https://doi.org/10.1007/s10490-023-09871-y

De Los Santos, J. A., Milla, N., Tuppal, C., & Labrague, L. (2022). Development and psychometric evaluation of the nurse behavior toward confirmed and suspected HIV/AIDS patients (NB-CSHAP) scale. *Frontiers of Nursing*, *9*(1), 103–113. https://doi.org/10.2478/fon-2022-0013

DeVellis, R. F., & Thorpe, C. T. (2022). *Scale development: Theory and applications* (5th ed.). SAGE. https://www.amazon.com/Scale-Development-Applications-Applied-Research/dp/154437934X

Elsman, E. B. M., Mokkink, L. B., Terwee, C. B., Beaton, D., Gagnier, J. J., Tricco, A. C., Baba, A., Butcher, N. J., Smith, M., Hofstetter, C., Lee Aiyegbusi, O., Berardi, A., Farmer, J., Haywood, K. L., Krause, K. R., Markham, S., Mayo-Wilson, E., Mehdipour, A., Ricketts, J., Szatmari, P., Touma, Z., moher, D., & Offringa, M. (2024). Guideline for reporting systematic reviews of outcome measurement instruments (OMIs): PRISMA-COSMIN for OMIs 2024. *Journal of Clinical Epidemiology, 173*, 111422. https://doi.org/10.1016/j.jclinepi.2024.111422

Gurung, D., Subedi, B., Kohrt, B. A., Wahid, S. S., Rai, S., Thornicroft, G., & Gronholm, P. C. (2025). Development of a mental health-related structural stigma measurement framework in the healthcare system setting: A modified Delphi study. *PLOS ONE, 20*(1), e0316999. https://doi.org/10.1371/journal.pone.0316999

Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2022). *A primer on partial least squares structural equation modeling (PLS-SEM)* (3rd ed.). SAGE Publications. https://uk.sagepub.com/en-gb/eur/a-primer-on-partial-least-squares-structural-equation-modeling-pls-sem/book270548

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning: With applications in R*. Springer US. https://doi.org/10.1007/978-1-0716-1418-1

JASP Team. (2025). *JASP* (Version 0.95.2) [Computer software]. https://jasp-stats.org

Kline, R. B. (2023). *Principles and practice of structural equation modeling* (5th ed.). The Guilford Press. https://www.guilford.com/books/Principles-and-Practice-of-Structural-Equation-Modeling/RexKline/9781462551910?srsltid=AfmBOor2eVifGKTSrjeXf8X7gAfSLHNfsvFBWTHQ5KQWz29Kv8RkSuIw

Koseoglu Ornek, O., Tabak, F., & Mete, B. (2020). Stigma in hospital: An examination of beliefs and attitudes towards HIV/AIDS patients, Istanbul. *AIDS Care, 32*(8), 1045–1051. https://doi.org/10.1080/09540121.2020.1769833

Loureiro, L., Santos, J., & Loureiro, C. (2025). Mental Health Continuum—Short Form: Confirmatory factor analysis (CFA) of competing models with adolescents from Portugal. *European Journal of Investigation in Health, Psychology and Education, 15*(4), 43. https://doi.org/10.3390/ejihpe15040043

Luz, P. M., Torres, T. S., Almeida-Brasil, C. C., Marins, L. M. S., Bezerra, D. R. B., Veloso, V. G., Grinsztejn, B., Harel, D., & Thombs, B. D. (2020). Translation and validation of the Short HIV Stigma Scale in Brazilian Portuguese. *Health and Quality of Life Outcomes, 18*(1), 322. https://doi.org/10.1186/s12955-020-01571-1

Ngcobo, S. J., Makhado, L., & Sehularo, L. A. (2024). Registered nurses' experiences regarding operational factors influencing the implementation of HIV care services in the mobile health clinics of eThekwini Municipality in KwaZulu-Natal. *BMC Nursing, 23*(1), 102. https://doi.org/10.1186/s12912-024-01764-9

Pearl, R. L., Li, Y., Groshon, L. C., Hernandez, M., Saunders, D., Sheynblyum, M., Driscoll, K. A., Gelfand, J. M., Manavalan, P., Montanez-Wiscovich, M., Pereira, D. B., Puhl, R. M., Wadden, T. A., Waxenberg, L. B., Westen, S. C., & Lou, X.-Y. (2024). Measuring internalized health-related stigma across health conditions: Development and validation of the I-HEARTS Scale. *BMC Medicine, 22*(1), 435. https://doi.org/10.1186/s12916-024-03661-z

Posit team. (2025). *RStudio: Integrated development environment for R* (Version 2025.9.0.387) [Computer software]. Posit Software, PBC. http://www.posit.co/

R Core Team. (2025). *R: A language and environment for statistical computing* (Version 4.5.1) [Computer software]. R Foundation for Statistical Computing. https://www.r-project.org/

Relf, M. V., Holzemer, W. L., Holt, L., Nyblade, L., & Ellis Caiola, C. (2021). A review of the state of the science of HIV and stigma: Context, conceptualization, measurement, interventions, gaps, and future priorities. *Journal of the Association of Nurses in AIDS Care, 32*(3), 392–407. https://doi.org/10.1097/JNC.0000000000000237

Roemer, E., Schuberth, F., & Henseler, J. (2021). HTMT2—An improved criterion for assessing discriminant validity in structural equation modeling. *Industrial Management & Data Systems, 121*(12), 2637–2650. https://doi.org/10.1108/IMDS-02-2021-0082

Rogers, P. (2024). Best practices for your confirmatory factor analysis: A JASP and lavaan tutorial. *Behavior Research Methods, 56*(7), 6634–6654. https://doi.org/10.3758/s13428-024-02375-7

S, S., & Mohanasundaram, T. (2024). Fit indices in structural equation modeling and confirmatory factor analysis: Reporting guidelines. *Asian Journal of Economics, Business and Accounting, 24*(7), 561–577. https://doi.org/10.9734/ajeba/2024/v24i71430

Teodorescu, V., & Obreja Brașoveanu, L. (2025). Assessing the validity of k-fold cross-validation for model selection: Evidence from bankruptcy prediction using random forest and XGBoost. *Computation, 13*(5), 127. https://doi.org/10.3390/computation13050127

United Nations. (2015). *Transforming our world: The 2030 Agenda for Sustainable Development*. https://sdgs.un.org/2030agenda

Wang, Y.-M., Hung, C.-H., Li, Y.-C., Ho, Y.-C., & Huang, C.-Y. (2025). Evaluating the construct validity of the health promotion literacy scale: A confirmatory factor analysis in Taiwan's university social responsibility context. *BMC Medical Education, 25*(1), 428. https://doi.org/10.1186/s12909-025-06992-4

World Health Organization. (2022). *Global health sector strategies on HIV, viral hepatitis and sexually transmitted infections for the period 2022–2030*. https://www.who.int/publications/i/item/9789240053779

World Health Organization. (2023). *WHO releases technical brief on reducing HIV-related stigma and discrimination in healthcare settings*. https://www.who.int/news/item/22-07-2024-who-releases-technical-brief-on-reducing-hiv-related-stigma-and-discrimination-in-healthcare-settings

Zhang, Y., Yang, X., Chai, X., Han, S., Zhang, L., Shao, Y., Ma, J., Li, K., & Wang, Z. (2024). Psychometric properties of stigma and discrimination measurement tools for persons living with HIV: A systematic review using the COSMIN methodology. *Systematic Reviews, 13*(1), 115. https://doi.org/10.1186/s13643-024-02535-y