Psychometric Evaluation of the Indonesian SF-12 Health Survey among Informal Female Workers: A CFA Approach

Dina Lusiana Setyowati[1], Nur Rohmah[2], Hanifa M Denny[3], Indah Fitri Astuti[4], Fatimah[5]

{dinalusiana@fkm.unmul.ac.id[1], nurrohmah@fkm.unmul.ac.id[2], hanifadenny@live.undip.ac.id[3], indahfitriastuti@unmul.ac.id[4] , fati007@brin.go.id[5] }

[1,2]Public Health Faculty, Mulawarman University, East Kalimantan, Indonesia
[3]Public Health Faculty, Universitas Diponegoro, Semarang, Indonesia
[4]Engineering Faculty, Mulawarman University, East Kalimantan, Indonesia
[5]Research Center for Applied Botany National Research and Innovation Agency, Indonesia

Abstract. The Short Form-12 Health Survey (SF-12) is a widely used tool for assessing health-related quality of life (HRQoL). Cultural adaptation and psychometric validation are essential for accuracy across diverse populations, particularly in marginalized or small-sample settings. This study examined the construct validity and internal consistency of the Indonesian SF-12 among informal female workers using a unidimensional model. A cross-sectional survey was conducted with 31 participants. Confirmatory Factor Analysis (CFA) using Maximum Likelihood Estimation in JASP assessed model fit (CFI, TLI, RMSEA, SRMR). SF-12 showed high reliability (Cronbach's alpha = 0.884; McDonald's omega = 0.878). Eleven of 12 items had significant factor loadings; SFa did not. Model fit indices and AVE (0.431) were below recommended cutoffs, indicating limited convergent validity. Despite strong reliability, the unidimensional structure showed inadequate fit, suggesting cultural misalignment. Future studies should explore alternative models, increase sample sizes, and adjust for cultural factors to improve psychometric robustness.

Keywords: SF-12 Health Survey; Psychometric Validation; Confirmatory Factor Analysis; Health-Related Quality of Life; Informal Workers

1 Introduction

The evaluation of health-related quality of life (HRQoL) has become the cornerstone of public health research, clinical evaluation, and policy planning. It allows for the measurement of the broader effects of illness, disability, and treatment on individual well-being, beyond traditional clinical indicators. Among the most widely utilized tools for this purpose is the Short Form-12 Health Survey (SF-12), a condensed version of SF-36. Designed to yield both Physical Component Summary (PCS) and Mental Component Summary (MCS) scores, SF-12 offers a time-efficient and user-friendly alternative, particularly valuable in settings constrained by time or resources [1]–[4]. In Indonesia, where informal employment accounts for a large portion of the labor force, HRQoL assessment is particularly important for women engaged in informal work who face socioeconomic vulnerability, limited healthcare access, and a dual burden of

economic and domestic responsibilities. These unique challenges underscore the urgency of conducting a localized validation of the SF-12 in this context.

Numerous validation studies have confirmed the SF-12's utility across diverse populations and health conditions, including those with chronic illnesses, such as stroke and diabetes [5], [6]. Cross-cultural applications in countries such as China, Greece, and Korea have confirmed their reliability and construct validity [3], [7], [8]. Despite these promising outcomes, researchers have cautioned against the assumption of universal applicability. Cultural nuances and differing interpretations of health-related concepts necessitate localized validation to ensure meaningful and accurate measurement [9], [10]. This need becomes more urgent in the Indonesian context, where informal female workers experience distinctive psychosocial and occupational stressors that may affect how they interpret and respond to health-related survey items.

However, the SF-12 has limitations. Its abbreviated format may reduce sensitivity in capturing subtle or domain-specific health issues such as post-stroke functional impairments or emotional disturbances in psychologically complex populations [5], [11], [12]. Additionally, ceiling effects, where scores cluster near the top end of the scale, can obscure health disparities, especially in high-functioning groups [13]. These limitations underscore the importance of continuous validation and context-specific adaptations. In particular, instruments validated in high-income or formal employment settings may not adequately capture the multidimensional health challenges of women in informal sectors in Indonesia.

Despite these concerns, the SF-12 remains favoured for its brevity and psychometric adequacy. Compared to more comprehensive instruments such as the SF-36 or EQ-5D, it balances the respondent burden with reliable outcome measurements [1], [14]. While the EQ-5D serves economic evaluation purposes, its limited dimensionality can hinder nuanced HRQoL assessments [15].

Confirmatory Factor Analysis (CFA) is the gold standard for testing an instrument's structural validity. It assesses the degree to which empirical data aligns with a theoretically derived factor structure. Benchmark fit indices include the Comparative Fit Index (CFI), Tucker-Lewis index (TLI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square Residual (SRMR). Conventionally, CFI and TLI values ≥ 0.95, and RMSEA ≤ 0.06, suggest an acceptable model fit [16], [17]. CFA also facilitates the detection of weak items, aiding model refinement and cultural adaptation [14], [18].

Given these advantages, CFA has become integral in validating instruments, such as the SF-12, particularly when used in distinct cultural or occupational contexts. Continuous reassessment of psychometric properties ensures that tools such as SF-12 remain theoretically grounded and practically effective. For informal female workers in Indonesia, this approach is essential to ensure that the instrument does not overlook context-specific stressors and cultural interpretations of health and quality of life.

This study contributes to this goal by applying a CFA to evaluate the Indonesian version of the SF-12 among informal female workers. By examining the factor structure and internal consistency, it assesses whether the instrument performs as expected in this unique demographic. This study employs modern CFA methodologies and open-source tools, offering new insights into adapting global health measures to localized contexts. Furthermore, by emphasizing the Indonesian setting, the study highlights how validation in marginalized groups can inform more equitable health policy and intervention strategies

2 Methods

This study employed a quantitative psychometric validation approach to assess the construct validity and internal consistency of the Indonesian version of the SF-12 Health Survey. The research design followed a cross-sectional structure, focusing on a one-time data collection process from the targeted population. The core analysis focused on evaluating a unidimensional factor structure representing general health-related quality of life (HRQoL), using Confirmatory Factor Analysis (CFA). All analyses were performed using the JASP software with Maximum Likelihood Estimation (MLE) procedures.

A total of 31 informal female workers participated in the study, each completing the 12-item SF-12 questionnaire. Participants were selected using a purposive sampling technique, targeting women engaged in informal occupations in Semarang, Indonesia. Inclusion criteria included being female, aged 18 years and above, currently working in informal employment, and willing to provide informed consent. While the SF-12 typically delineates physical and mental health domains, this study tested a single latent HRQoL factor to determine whether a unified construct could sufficiently account for item responses.

Model evaluation began with an unconstrained baseline, followed by a specified one-factor CFA model. The assessment relied on standard thresholds for model fit, including CFI and TLI ≥ 0.95, RMSEA ≤ 0.06, and SRMR ≤ 0.08, as recommended in psychometric guidelines [16], [17]. These indices were used as the interpretive benchmarks to determine whether the unidimensional model adequately captured the HRQoL construct in the target population.

The reliability and item-performance of the SF-12 were evaluated by examining factor loadings, item-total correlations, and Cronbach's alpha values. Standardized factor loadings ≥ 0.30 were considered acceptable, while Cronbach's alpha values above 0.70 were interpreted as adequate internal consistency. "Alpha if item deleted" analyses were conducted to determine the contribution of each item to the overall reliability.

Descriptive statistics of item means and standard deviations were examined to detect distributional issues such as potential ceiling effects [13]. Residual variances were also checked to confirm item-level uniqueness, with statistical significance indicating that unexplained variance remained across items.

Convergent validity was assessed through the Average Variance Extracted (AVE). An AVE value ≥ 0.50 was interpreted as evidence of acceptable convergent validity [18]. This suggests moderate convergent validity and underscores the need for item refinement or alternative factor modeling. Values below this threshold suggested the need for item refinement or reconsideration of factor structures.

2.1 MLE Assumptions and Sample Considerations

MLE assumes multivariate normality, typically assessed using skewness and kurtosis metrics [19], [20]. If assumptions are violated, researchers should consider transformations or robust estimation alternatives [21].

The sample size is another critical limitation. With only 31 participants, parameter estimates may lack stability and fit indices may be biased [20], [22]. While some CFA models tolerate small samples under strict conditions, a sample size of at least 200 is typically preferred for a robust estimation [23]. In small-sample settings, Bayesian approaches offer a more stable alternative [19]. In this study, the limited sample size reflects the challenges of accessing informal female worker populations in Indonesia, a group often underrepresented in research

due to their unstable work schedules and socioeconomic constraints. This contextual limitation should be considered when interpreting findings.

Model specification should be driven by strong theoretical grounding and informed by exploratory findings or conceptual frameworks to minimize overfitting [24]. Model identification principles, such as ensuring more observed variables than parameters, must also be adhered to for valid inferences [25]. Finally, the recommended fit thresholds (CFI, TLI > 0.90; RMSEA < 0.06; SRMR < 0.08) served as a reference point for model evaluation [26], [27]. Given the small sample size of this study, caution is advised when interpreting the findings. Nonetheless, this analysis provides a foundational assessment for future research involving larger, more representative samples. Future studies should expand recruitment strategies, collaborate with local women's associations, and employ mixed-method approaches to strengthen both quantitative reliability and cultural adaptation of the SF-12.

3 Results

This section presents the empirical findings from a Confirmatory Factor Analysis (CFA) conducted to evaluate the psychometric validity and internal consistency of the Indonesian version of the SF-12 Health Survey. The results were organized across five domains: participant demographics, reliability analysis, model fit evaluation, factor loadings, and convergent validity through Average Variance Extracted (AVE).

3.1 Demographics Characteristics

Table 1 displays the demographic profile of the 31 female informal workers recruited in Semarang. A majority (54.84%) were between the ages of 26 and 45, while 41.93% were over 45, and only one participant (3.23%) was younger than 25. Most respondents were married (54.84%), followed by widowed (35.48%) and unmarried individuals (9.68%).

Regarding educational background, 67.74% had completed senior high school, 29.03% had tertiary education, and only 3.23% had junior high education. In terms of work experience, 54.84% had been in their jobs for over seven years, 32.26% for four to six years, and 12.90% for one to three years. These demographic distributions are consistent with previous literature reporting prolonged occupational and psychosocial stress among mid-to-late-career informal female workers in Indonesia [28]–[30].

Table 1. Demographic Characteristics of Informal Female Worker Respondents

Characteristics	n	Percentage (%)
Age (Year Old)		
<25	1	3.23
26-45	17	54.84
>45	13	41.93
Marital Status		
Unmarried	3	9.68
Married	17	54.84
Widowed	11	35.48
Educational Level		
Primary School		
Junior High School	1	3.23
Senior High School	21	67.74
College	9	29.03
Tenure (Years)		

Characteristics	n	Percentage (%)
1-3	4	12.90
4-6	10	32.26
≥7	17	54.84

Source: *Primary data*

3.2 Reliability Analysis

Internal consistency was assessed using Cronbach's Alpha. The overall scale demonstrated high reliability with Alpha = 0.884, surpassing the 0.80 threshold for strong consistency. Item-total correlations were above 0.50 for ten items, except SFa (0.293) and SFe (0.462), with SFa falling below the 0.30 cutoff [31]. "Alpha-if-item-deleted" analysis identified SFB and SFD as having the most favorable impact on overall reliability.

Table 2. Reliability test results with Cronbach's Alpha

	mean	Std. deviation	Item-rest-correlation	If item dropped Cronbach's a
SFA	4.06	.854	.293	.890
SFB	2.55	.675	.765	.869
SFC	2.52	.724	.643	.873
SFD	3.94	.892	.769	.865
SFE	4.03	1.016	.462	.882
SFF	4.13	1.258	.752	.864
SFG	4.10	.978	.575	.876
SFH	4.03	.983	.626	.873
SFI	3.61	1.174	.618	.874
SFJ	3.97	.983	.546	.877
SFK	4.03	.795	.543	.877
SFL	4.13	.763	.560	.877

Source: *Primary data*

3.3 Model Fit & Fit Indices

Table 3 summarizes the CFA model fit indices derived via JASP. The unidimensional model failed to achieve acceptable fit. The CFI (0.723) and TLI (0.662) were well below the 0.90 benchmark. Similarly, the RMSEA (0.176) and SRMR (0.102) exceeded the maximum recommended cutoffs of 0.06 and 0.08, respectively [32], [33]. These metrics suggest the hypothesized model does not adequately represent the underlying data structure.

Table 3. Model Fit Indices of the CFA for the Indonesian SF-12

Fit Indices	Value
Comparative Fit Index (CFI)	0.723
Tucker-Lewis Index (TLI)	0.662
Standardiized root mean square residual (SRMR)	0.102
Root mean square error of approximation (RMSEA)	0.176

Source: *Primary data*

3.4 Factor Loadings

As shown in Table 4, eleven out of twelve SF-12 items exhibited statistically significant factor loadings ($p < .01$). SFf had the strongest loading ($\lambda = 1.017$), followed by SFi ($\lambda = 0.800$) and SFd ($\lambda = 0.754$). SFa did not reach significance ($\lambda = 0.268$, $p = 0.080$), implying limited

contribution to the overall HRQoL construct. All other items fell within acceptable confidence intervals, further validating their alignment with the latent factor.

Table 4. Standardized Factor Loadings

Factor loadings

| Factor	Indicator	Estimate	Std. Error	z-value	p	95% Confidence Interval	
						Lower	Upper
SF	SFa	0.268	0.153	1.752	0.080	-0.032	0.568
	SFb	0.563	0.098	5.720	< .001	0.370	0.756
	SFc	0.471	0.118	4.005	< .001	0.241	0.702
	SFd	0.754	0.129	5.829	< .001	0.500	1.007
	SFe	0.546	0.172	3.165	0.002	0.208	0.884
	SFf	1.017	0.186	5.456	< .001	0.652	1.382
	SFg	0.534	0.168	3.185	0.001	0.206	0.863
	SFh	0.624	0.162	3.856	< .001	0.307	0.940
	SFi	0.800	0.187	4.269	< .001	0.433	1.168
	SFj	0.528	0.167	3.161	0.002	0.201	0.855
	SFk	0.382	0.139	2.756	0.006	0.110	0.654
	SFl	0.409	0.130	3.145	0.002	0.154	0.664

Source: *Primary data*

3.5 Reliability and AVE

Figure 1 and 2 shows the assess convergent validity, the Average Variance Extracted (AVE) was calculated at 0.431, slightly below the 0.50 benchmark [18]. This indicates that less than half of the item variance is accounted for by the latent HRQoL factor. While not invalidating, the result signals moderate convergent validity and calls for potential revision of item content or consideration of alternative model structures.
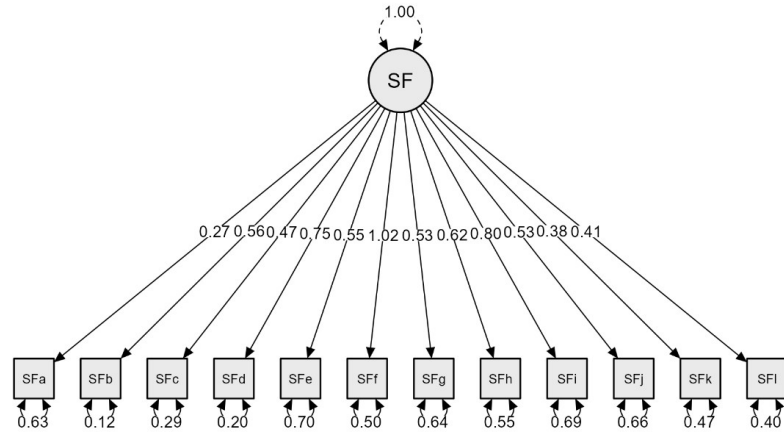
Fig. 1. Confirmatory Factor Analysis (CFA) Model Framework for the Indonesian SF-12
Source: *Primary data*
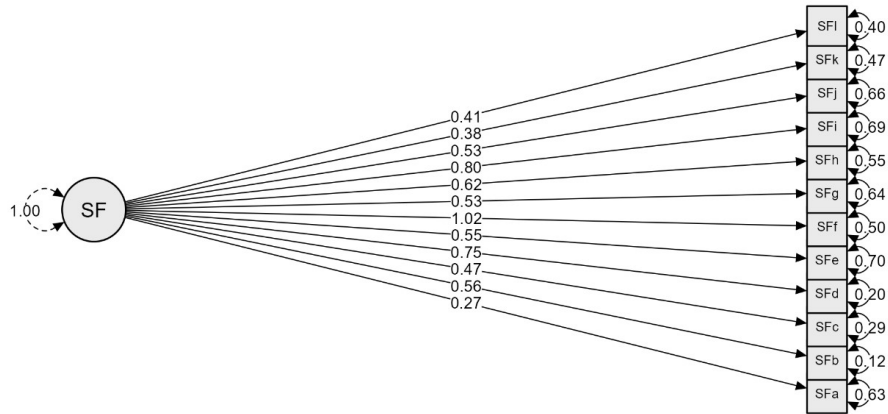
Fig. 2. Reliability Indices and Variance Explained for the Indonesian SF-12
Source: *Primary data*

4 Discussion

This study examined the psychometric properties of the Indonesian version of the SF-12 Health Survey with a focus on factorial validity and internal consistency among informal female workers. By combining quantitative metrics with contextual interpretation, the discussion outlines both the limitations of current modeling and opportunities for future scale adaptation. The findings are contextualized within the Indonesian setting, where female informal workers such as street vendors, domestic helpers, and small traders face limited access to healthcare, economic precarity, and the dual burden of paid work and household responsibilities. These contextual realities are crucial to understanding how this population interprets and responds to HRQoL items.

4.1 Interpretation of Poor Model Fit Despite High Internal Reliability

A key paradox in this study was the discordance between high internal consistency and poor model fit. While reliability indices (Cronbach's Alpha = 0.884) indicate strong internal coherence, the CFA model fit indices (CFI = 0.723; TLI = 0.662; RMSEA = 0.176; SRMR = 0.102) fell well below accepted thresholds. This inconsistency raises concerns about the appropriateness of applying a unidimensional HRQoL structure within this demographic.

These findings suggest that although items may share common variance, they do not accurately represent a single latent construct indicating a potential misspecification of the model or contextual misalignment [34]. Contributing factors may include the small sample size, which limits statistical power [35], and culturally specific interpretations of health-related questions [36]. For example, items such as SFa ("general health perception") may be understood differently by Indonesian informal workers, who often prioritize immediate functional ability over abstract notions of wellness. While a single-factor model may be theoretically defensible in some populations [37], the evidence here advocates for exploring alternative structures, such as bifactor or two-dimensional models that differentiate physical and mental health.

4.2 Rationale for Model Modifications

Given the inadequate fit, CFA model refinements should be considered. Modifications may involve correlating residuals, excluding poorly performing items, or restructuring latent constructs [38]. However, such adjustments must be theory-driven to avoid overfitting and loss of external validity [39]. Despite model limitations, the scale retained substantive item coherence, with 11 out of 12 items showing significant loadings. The exception was item SFa ("general health perception"), which displayed a low item-total correlation (0.293) and non-significant factor loading ($\lambda = 0.268$). This suggests that respondents may have interpreted the question in culturally specific ways, such as equating "health" primarily with the absence of visible illness rather than overall well-being. These metrics, coupled with the suboptimal AVE (0.431), highlight the need for item-level revisions or potential removal of misaligned items [18].

4.3 Cultural Adaptation Challenges

Adapting health instruments like the SF-12 across cultural contexts presents inherent challenges. Local semantics, values, and health perceptions can significantly alter the interpretation of survey items [40], [41]. Translation alone may not suffice to achieve measurement equivalence, particularly in populations with varying literacy levels and occupational stressors. For informal female workers in Indonesia, limited formal education and the dominance of pragmatic health concerns (e.g., ability to continue working despite illness) may lead to different interpretations compared with populations in high-income settings. Effective adaptation strategies include participatory approaches involving local stakeholders, iterative refinement through cognitive interviews, and pre-testing to identify ambiguities [34], [42]. These steps ensure that the instrument retains psychometric integrity while becoming contextually relevant.

4.4 Demographic and Contextual Insights

The demographic characteristics of the sample predominantly middle-aged, married or widowed women with extensive work experience underscore the influence of occupational fatigue and domestic burdens on HRQoL perceptions [30]. These factors, compounded by socioeconomic constraints, may shape unique response patterns that deviate from the assumptions embedded in Western-developed tools. For instance, balancing caregiving duties with precarious informal work may create unique health burdens that are not fully captured by standard SF-12 items. In this context, administering the SF-12 without modification risks underrepresenting psychosocial stressors. Enhancing construct representation may require integrating locally meaningful health indicators or tailoring item phrasing to match respondents' lived experiences [29].

4.5 Implications and Recommendations

Practically, the SF-12 showed adequate reliability for use as a quick HRQoL screening tool. However, the poor structural fit and modest AVE values caution against using the total score for analytical purposes without further validation.

Theoretically, this study affirms the need for multidimensional models in culturally diverse settings. Future research should explore bifactor or domain-specific CFA structures and assess improvements in model fit and interpretability. Additionally, qualitative methods such as cognitive debriefing can uncover item misinterpretations and cultural biases.

Scaling up the sample size is critical for statistical robustness and generalizability. In addition, future studies should test the instrument's invariance across different subgroups of

Indonesian women, including variations in age, region, and occupation, to ensure broader applicability. Incorporating constructs like work fatigue, emotional burden, and social support could enhance content validity and provide a more nuanced understanding of HRQoL among informal workers. These steps will ultimately contribute to more equitable and culturally sensitive health assessment frameworks.

5 Conclusion

This study evaluated the psychometric characteristics of the Indonesian version of the SF-12 Health Survey among informal female workers, employing Confirmatory Factor Analysis (CFA) to assess construct validity and internal consistency. The instrument demonstrated high reliability, as evidenced by Cronbach's Alpha (0.884) and McDonald's Omega (0.878), suggesting that the SF-12 items consistently measure general health-related quality of life (HRQoL) within this group. However, reliability alone is not sufficient to validate the structure of the instrument without supportive model fit evidence. Indonesian context, particularly among women working in informal sectors with socioeconomic vulnerability and dual work family burdens, makes this validation critical for ensuring measurement accuracy.

However, the structural validity of the instrument was not supported by the CFA results. Model fit indices specifically the Comparative Fit Index (CFI = 0.723), Tucker-Lewis Index (TLI = 0.662), Root Mean Square Error of Approximation (RMSEA = 0.176), and Standardized Root Mean Square Residual (SRMR = 0.102) failed to meet acceptable benchmarks. Additionally, only 11 out of 12 items significantly loaded onto the latent factor, and the Average Variance Extracted (AVE) was below the recommended threshold (0.431 < 0.50), indicating inadequate convergent validity. Item SFa ("general health perception") was non-significant, reflecting possible cultural misalignment in how respondents conceptualize overall health, which emphasizes functional ability rather than abstract self-rated health.

These findings suggest that a unidimensional structure may not sufficiently capture the multifaceted nature of HRQoL in this population. This discrepancy underscores the need to rethink the structural assumptions underlying the SF-12 in cross-cultural contexts. Relying on standardized constructs without context-specific validation risks misrepresenting the lived health experiences of target populations. In particular, Indonesian female informal workers may interpret items through the lens of economic survival, caregiving responsibilities, and fatigue, which standard HRQoL measures often fail to address.

Future research should investigate the applicability of multidimensional or bifactor models that distinguish between physical and mental health dimensions. Testing model invariance and structural validity across different subgroups may also enhance generalizability. Incorporating culturally adapted items, qualitative insights, and larger sample sizes will further strengthen the instrument's contextual validity and measurement precision. Additionally, mixed-methods approaches, including cognitive interviewing, can provide deeper understanding of how respondents interpret survey items. Special attention should be given to items like SFa, which may require rewording or contextual adaptation to capture culturally relevant aspects of health among Indonesian women.

This study contributes valuable evidence supporting the importance of localized psychometric evaluation. It lays a methodological foundation for future efforts to enhance the relevance and fairness of global health instruments, particularly in underrepresented labor populations. By focusing on informal female workers in Indonesia, this study highlights the necessity of adapting global measures to reflect local cultural values, work conditions, and gendered health experiences.

Acknowledgement

Funding

References

[1]   A. Montazeri, M. Vahdaninia, S. J. Mousavi, and S. Omidvari, "The Iranian version of 12-item Short Form Health Survey (SF-12): factor structure, internal consistency and construct validity," *BMC Public Health*, vol. 9, no. 1, p. 341, Dec. 2009, doi: 10.1186/1471-2458-9-341.

[2]   J. CR, A. HR, S. Sridhar, K. Akhila, S. Sridhar, and R. Kumawat, "A Randomized Control Study to Evaluate the Role of Herbal Immunomodulators in Boosting the Immunity and Overall Health of Healthcare Workers in Covid-19 Wards: An Exploratory, Feedback Clinical Study," *Asian J. Pharm. Clin. Res.*, pp. 138–142, 2021, doi: 10.22159/ajpcr.2021.v14i8.42035.

[3]   S. Kim, M. Jo, J. Ahn, M. Ock, S. Shin, and J. Park, "Assessment of psychometric properties of the Korean SF-12 v2 in the general population," *BMC Public Health*, vol. 14, no. 1, p. 1086, Dec. 2014, doi: 10.1186/1471-2458-14-1086.

[4]   M. Obtel, K. El Rhazi, S. Elhold, M. Benjelloune, L. Gnatiuc, and C. Nejjari, "Cross-cultural adaptation of the 12-Item Short-Form survey instrument in a Moroccan representative Survey," *South. African J. Epidemiol. Infect.*, vol. 28, no. 3, pp. 166–171, Jan. 2013, doi: 10.1080/10158782.2013.11441540.

[5]   O. C. Okonkwo, D. L. Roth, L. Pulley, and G. Howard, "Confirmatory factor analysis of the validity of the SF-12 for persons with and without a history of stroke," *Qual. Life Res.*, vol. 19, no. 9, pp. 1323–1331, Nov. 2010, doi: 10.1007/s11136-010-9691-8.

[6]   A. H. Al-Shehri, A. Z. Taha, A. Bahnassy, and M. Salah, "Health-Related Quality of Life in Type 2 Diabetic Patients," *Ann. Saudi Med.*, vol. 28, no. 5, pp. 352–360, 2008, doi: 10.5144/0256-4947.2008.352.

[7]   C. L. K. Lam, E. Y. Y. Tse, and B. Gandek, "Is the standard SF-12 Health Survey valid and equivalent for a Chinese population?," *Qual. Life Res.*, vol. 14, no. 2, pp. 539–547, Mar. 2005, doi: 10.1007/s11136-004-0704-3.

[8]   N. Kontodimopoulos, E. Pappa, D. Niakas, and Y. Tountas, "Validity of SF-12 summary scores in a Greek general population," *Health Qual. Life Outcomes*, vol. 5,

no. 1, p. 55, Dec. 2007, doi: 10.1186/1477-7525-5-55.

[9]   N. Wang, F. Ren, and X. Zhou, "Factor Structure and Psychometric Properties of the Body Perception Questionnaire–Short Form (BPQ-SF) Among Chinese College Students," *Front. Psychol.*, vol. 11, 2020, doi: 10.3389/fpsyg.2020.01355.

[10]   R. Shah *et al.*, "An Evaluation of the Psychometric Properties of the Sf-12v2 Health Survey Among Adults With Hemophilia," *Health Qual. Life Outcomes*, vol. 16, no. 1, 2018, doi: 10.1186/s12955-018-1059-8.

[11]   P. Menn, N. Weber, and R. Holle, "Health-Related Quality of Life in Patients With Severe COPD Hospitalized for Exacerbations - Comparing EQ-5D, SF-12 and SGRQ," *Health Qual. Life Outcomes*, vol. 8, no. 1, p. 39, 2010, doi: 10.1186/1477-7525-8-39.

[12]   W. J. Rejeski, E. H. Ip, A. P. Marsh, and R. Barnard, "Development and Validation of a Video-Animated Tool for Assessing Mobility," *Journals Gerontol. Ser. A*, vol. 65A, no. 6, pp. 664–671, 2010, doi: 10.1093/gerona/glq055.

[13]   U. Jakobsson, A. Westergren, S. Lindskov, and P. Hagell, "Construct validity of the SF-12 in three different samples," *J. Eval. Clin. Pract.*, vol. 18, no. 3, pp. 560–566, Jun. 2012, doi: 10.1111/j.1365-2753.2010.01623.x.

[14]   D. Y. T. Fong, J. Y. H. Wong, E. P. H. Choi, K. F. Lam, and C. Kwok, "The English and Chinese language versions of the Short Form 12-item Health Survey are equivalent," *Health Qual. Life Outcomes*, vol. 19, no. 1, p. 14, Dec. 2021, doi: 10.1186/s12955-020-01653-0.

[15]   B. Gandek, E. M. Roos, P. D. Franklin, and J. E. Ware, "Item selection for 12-item short forms of the Knee injury and Osteoarthritis Outcome Score (KOOS-12) and Hip disability and Osteoarthritis Outcome Score (HOOS-12)," *Osteoarthr. Cartil.*, vol. 27, no. 5, pp. 746–753, May 2019, doi: 10.1016/j.joca.2018.11.011.

[16]   P. A. Lowe, "Cross-National Comparisons between Canadian and US Higher Education Students on a New, Brief, Multidimensional Measure of Test Anxiety," *J. Psychoeduc. Assess.*, vol. 39, no. 6, pp. 665–679, Sep. 2021, doi: 10.1177/07342829211016933.

[17]   D. T. Quintana, M. P. Casanova, A. C. Cady, and R. T. Baker, "Assessing the Structural Validity of the Knee Injury and Osteoarthritis Outcome Score Scale," *Healthcare*, vol. 12, no. 4, p. 414, Feb. 2024, doi: 10.3390/healthcare12040414.

[18]   Y. Lim, "Psychometric Properties of the Mental Health Continuum-Short Form (MHC-SF): A Study With Individuals With Schizophrenia Living in the Community," *Psychiatry Investig.*, vol. 19, no. 12, pp. 1021–1026, Dec. 2022, doi: 10.30773/pi.2022.0155.

[19]   D. Harrington, "Assessing Confirmatory Factor Analysis," in *Confirmatory Factor Analysis*, Oxford University Press, 2008, pp. 50–77. doi: 10.1093/acprof:oso/9780195339888.003.0004.

[20]   J. C. F. de Winter and D. Dodou, "Factor recovery by principal axis factoring and maximum likelihood factor analysis as a function of factor pattern and sample size," *J. Appl. Stat.*, vol. 39, no. 4, pp. 695–710, Apr. 2012, doi: 10.1080/02664763.2011.610445.

[21]   A. I. Cuesta-Vargas *et al.*, "Psychometric properties of the QuickPIPER: a shortened version of the PIPER Fatigue scale," *Eur. J. Cancer Care (Engl).*, vol. 22, no. 2, pp. 245–252, Mar. 2013, doi: 10.1111/ecc.12022.

[22]   E. A. Erekson *et al.*, "The Vulvovaginal Symptoms Questionnaire," *Menopause*, vol. 20, no. 9, pp. 973–979, Sep. 2013, doi: 10.1097/GME.0b013e318282600b.

[23]   J.-F. Wu, "Examining Chen and Starosta's Model of Intercultural Sensitivity in the Taiwanese Cultural Context," *Int. J. Mod. Educ. Comput. Sci.*, vol. 7, no. 6, pp. 1–8, Jun. 2015, doi: 10.5815/ijmecs.2015.06.01.

[24]   S. Rajaa, Y. Krishnamoorthy, and J. Ramakrishnan, "Psychometric properties of the Tamil version of Perceived Stress Scale among diabetes mellitus patients in Puducherry, South India," *J. Fam. Med. Prim. Care*, vol. 11, no. 8, pp. 4688–4693, Aug. 2022, doi: 10.4103/jfmpc.jfmpc_2346_21.

[25]   S. Ghanbari-Homaie, M. Asghari Jafarabadi, S. Hasani, and M. Mirghafourvand, "Psychometric Properties of the Persian Version of Pregnancy Symptoms Inventory (PSI)," *Clin. Nurs. Res.*, vol. 31, no. 5, pp. 960–967, Jun. 2022, doi: 10.1177/10547738211067319.

[26]   H. Stuart, N. Sartorius, and T. Liinamaa, "The images of psychiatry scale: development, factor structure, and reliability," *BMC Psychiatry*, vol. 14, no. 1, p. 337, Dec. 2014, doi: 10.1186/s12888-014-0337-1.

[27]   L. N. Jespersen, A. Andersen, P. Due, A. K. Ersbøll, M. T. Damsgaard, and S. I. Michelsen, "From questionnaire to practical application: A quality of life profile for populations with diverse disabilities," *Scand. J. Public Health*, vol. 50, no. 3, pp. 362–370, May 2022, doi: 10.1177/1403494820981437.

[28]   M. R. Frone, D. Reis, and C. Ottenstein, "A German Version of the Three-Dimensional Work Fatigue Inventory (3<scp>D</Scp>-<scp>WFI</Scp>): Factor Structure, Internal Consistency, and Correlates," *Stress Heal.*, vol. 34, no. 5, pp. 674–680, 2018, doi: 10.1002/smi.2828.

[29]   A. K. Danielsen, H. Pommergaard, J. Burcharth, E. Angenete, and J. Rosenberg, "Translation of Questionnaires Measuring Health Related Quality of Life Is Not Standardized: A Literature Based Research Study," *PLoS One*, vol. 10, no. 5, p. e0127050, 2015, doi: 10.1371/journal.pone.0127050.

[30]   A. Bazazan, Y. Rasoulzadeh, I. Dianat, A. Safaiyan, and Z. Mombeini, "Occupational fatigue and mental health complaints among 8-hour shift workers of petrochemical industries in Iran," *Work*, vol. 62, no. 2, pp. 309–317, Feb. 2019, doi: 10.3233/WOR-192865.

[31]   A. Escobar, J. M. Quintana, A. Bilbao, I. Aróstegui, I. Lafuente, and I. Vidaurreta, "Responsiveness and clinically important differences for the WOMAC and SF-36 after total knee replacement," *Osteoarthr. Cartil.*, vol. 15, no. 3, pp. 273–280, Mar. 2007, doi: 10.1016/j.joca.2006.09.001.

[32]   Z. Keshavarz, M. Simbar, A. Ramezankhani, and H. A. Majd, "An inventory for assessment of female workers' health promotion behaviour based on the integrated model of planned behaviour and self-efficacy," *East. Mediterr. Heal. J.*, vol. 19, no. 6, pp. 561–569, Jun. 2013, doi: 10.26719/2013.19.6.561.

[33]   H. Kim, B. Ku, J. Y. Kim, Y.-J. Park, and Y.-B. Park, "Confirmatory and Exploratory Factor Analysis for Validating the Phlegm Pattern Questionnaire for Healthy Subjects," *Evidence-Based Complement. Altern. Med.*, vol. 2016, no. 1, Jan. 2016, doi: 10.1155/2016/2696019.

[34]   F. Ozcan and A. Meydan, "Zero Waste Attitude Scale Development Study," *J. Educ. Sci. Environ. Heal.*, pp. 69–89, Jan. 2024, doi: 10.55549/jeseh.1419382.

[35]   Y. Zhang *et al.*, "Development and Validation of a Chinese Parental Health Literacy Questionnaire for Caregivers of Children from 0 to 3 Years Old." Jun. 25, 2019. doi:

10.21203/rs.2.10641/v1.

[36] V. Swami, C. Maïano, A. Furnham, and C. Robinson, "The intuitive eating scale-2: re-evaluating its factor structure using a bifactor exploratory structural equation modelling framework," *Eat. Weight Disord. - Stud. Anorexia, Bulim. Obes.*, vol. 27, no. 4, pp. 1349–1357, May 2022, doi: 10.1007/s40519-021-01271-9.

[37] S. Ng, "Validation of the 10-item Chinese perceived stress scale in elderly service workers: one-factor versus two-factor structure," *BMC Psychol.*, vol. 1, no. 1, p. 9, Dec. 2013, doi: 10.1186/2050-7283-1-9.

[38] M. J. Annear, C. Toye, K.-E. J. Elliott, F. McInerney, C. Eccleston, and A. Robinson, "Dementia knowledge assessment scale (DKAS): confirmatory factor analysis and comparative subscale scores among an international cohort," *BMC Geriatr.*, vol. 17, no. 1, p. 168, Dec. 2017, doi: 10.1186/s12877-017-0552-y.

[39] O. Ozkok, M. J. Zyphur, A. Barsky, M. Theilacker, M. B. Donnellan, and F. L. Oswald, "Modeling Measurement as a Sequential Process: Autoregressive Confirmatory Factor Analysis (AR-CFA)," *Front. Psychol.*, vol. 10, 2019, doi: 10.3389/fpsyg.2019.02108.

[40] Y. Liao *et al.*, "Development and Psychometric Evaluation of an Instrument to Assess Physical Activity Participation Behavior Among Chinese Patients with Chronic Obstructive Pulmonary Disease," *Int. J. Chron. Obstruct. Pulmon. Dis.*, vol. Volume 20, pp. 1321–1334, Apr. 2025, doi: 10.2147/COPD.S511061.

[41] A. Campo-Arias, A. F. T. Otálvaro, I. Álvarez-Solorza, and C. A. Cassiani-Miranda, "Confirmatory Factor Analysis, Internal Consistency, Gender Differential Item Functioning and Discriminant Validity of the Fear of COVID-5 Scale Amidst Emerging Adult University Students in Mexico," *Omega - J. Death Dying*, vol. 87, no. 1, pp. 4–19, 2021, doi: 10.1177/00302228211016216.

[42] E. Esen, "The My Children's Future Scale : Construct validity, measurement invariance, and reliability in a Turkish sample," *Aust. J. Career Dev.*, vol. 29, no. 3, pp. 185–195, Oct. 2020, doi: 10.1177/1038416220941853.