Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Translation and Validation of a Theoretically-Informed Questionnaire to Assess the Acceptability of Healthcare Interventions among Filipino Older Adults

Angely P. Garcia

apgarcia@up.edu.ph

Institute on Aging, National Institutes of Health, University of the Philippines Manila
https://orcid.org/0000-0003-1022-1862

Zypher Jude G. Regencia

Department of Sociology and Behavioral Sciences, College of Liberal Arts, De La Salle University
https://orcid.org/0000-0002-3383-6336

Olivia T. Sison

Institute of Clinical Epidemiology, National Institutes of Health, University of the Philippines Manila
https://orcid.org/0000-0002-0929-3052

Shelley Ann F. de la Vega

College of Medicine, University of the Philippines Manila https://orcid.org/0000-0003-0350-0812

Research Article

Keywords: acceptability, cross-cultural adaptation, older adults, Philippines, psychometric validation, Theoretical Framework of Acceptability

Posted Date: June 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9988030/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Abstract

The Theoretical Framework of Acceptability (TFA) offers a comprehensive seven-construct model for assessing intervention acceptability. No validated Filipino-language TFA tool existed for older adults prior to this study. A three-phase mixed-methods study was conducted: translation and adaptation, cognitive interviews (n=12), and pilot psychometric validation (n=80 older adults, mean age 69.3 years, 75% female) from Barangay UP Campus, Diliman, Quezon City, Philippines. Exploratory factor analysis (EFA), Cronbach's alpha, intraclass correlation coefficient (ICC), and content validity indices were computed. Data was collected between August and September 2025. Content validity was excellent (S-CVI/Ave = 0.97). EFA yielded a two-factor structure, explaining 48.4% of the variance, with eigenvalues of 3.38 and 1.46: Factor 1, "Perceived Value & Ethical-Appraisal," and Factor 2, "Practical Engagement & Comprehension." Internal consistency was acceptable (Cronbach's α = 0.73). Test-retest reliability was poor (ICC = −0.06, n = 10) and should be interpreted with caution given the small sample size. The Filipino TFA questionnaire demonstrates good content validity and acceptable internal consistency, providing a culturally appropriate tool for assessing health intervention acceptability among Filipino older adults. Further validation with larger samples is recommended to establish temporal stability.

Introduction

Global population aging represents one of the most significant demographic transformations of the 21st century, with the number of adults aged 60 years and older projected to reach 2.1 billion worldwide by 2050 (Grinin *et al*. 2023). The Philippines is no exception from this rapid demographic transition, with older adults accounting for 9.6% of the nation's 113.7 million people (He *et al*. 2022). Filipino older adults face a substantial burden of non-communicable diseases, including diabetes, hypertension, and cardiovascular disease, yet utilization of preventive health programs and adherence to therapeutic interventions remain suboptimal (Belarga *et al*. 2022; Castillo 2024). Understanding and enhancing intervention acceptability is, therefore, critical to improving engagement with health services in this rapidly aging population.

Acceptability has been recognized as a key implementation outcome in health services research, yet it remains under-assessed and poorly measured, with studies frequently relying on behavioral proxies such as dropout rates rather than validated instruments that directly assess recipients' perceptions (Sekhon *et al.* 2017). Commonly used acceptability tools include the acceptability intervention measures (AIM), with test-retest reliability coefficients ranging from 0.73 to 0.88; the abbreviated acceptability rating profile; the client satisfaction questionnaire-8 (CSQ-8); and qualitative methods such as focus group discussions and interviews. Validated, theory-informed instruments are needed to standardize acceptability measurement and enable cross-study comparisons (Sekhon *et al.* 2017; Nuñez *et al.* 2022; Sekhon *et al.* 2022).

One of the most cited, adapted, and more recently developed frameworks on acceptability, based on a systematic review of the health services and behavioral sciences literature, is the Theoretical Framework

of Acceptability (TFA) (Sekhon *et al.* 2017; Sekhon *et al.* 2022). It consists of seven component constructs: Affective Attitude, Burden, Ethicality, Intervention Coherence, Opportunity Costs, Perceived Effectiveness, and Self-Efficacy. The generic English TFA questionnaire is an adaptable tool that researchers and clinicians can use to assess the acceptability of interventions in various healthcare settings (Sekhon *et al.* 2022). Cross-cultural adaptations of the TFA questionnaire have been developed in Danish (Bekker-Jensen *et al.* 2026) and other languages (e.g., Swedish, Spanish), with consistent reporting of difficulties with ethicality and self-efficacy constructs, suggesting these constructs may be particularly sensitive to cultural and linguistic context (Rivera *et al.* 2025; Samuelsson *et al.* 2025).

The Philippines is a linguistically diverse nation, and older Filipino adults often have modest educational backgrounds, which can pose comprehension barriers to abstract or technical terminology (Judan-Ruiz *et al.* 2020; Torres and Macindo 2020). Moreover, Filipino culture is characterized by collectivist values in which individual decisions and moral judgments are embedded within family and community context (Vasquez *et al.* 2001), with important implications for the translation and validation of acceptability measures. Prior to the present study, no validated Filipino-language TFA questionnaire existed, limiting the capacity of Philippine health researchers to systematically assess intervention acceptability using a theoretically grounded, culturally appropriate tool, a gap consistent with broader challenges in Filipino instrument development (Judan-Ruiz *et al.* 2020; Torres and Macindo 2020; Tan *et al.* 2025).

Against this backdrop, this study aimed to translate, culturally adapt, and validate the TFA questionnaire for Filipino older adults through forward-backward translation, expert content validity review, cognitive interviews, and pilot psychometric testing, including exploratory factor analysis, internal consistency, and test-retest reliability, following established cross-cultural adaptation methodologies widely applied in the Philippine context (Judan-Ruiz *et al.* 2020; Torres and Macindo 2020; Nuñez *et al.* 2022; Tan *et al.* 2025).

Methods

Study Design

This study employed a mixed-methods, three-phase sequential design following established guidelines for cross-cultural adaptation of self-report measures (Beaton *et al.* 2000; Rupareliya and Shukla 2020) and the Consensus-based Standards for the selection of health Measurement Instruments (COSMIN) checklist (Mokkink *et al.* 2016). Phase 1 consisted of forward and backward translation and expert committee review to assess content validity. Phase 2 involved cognitive interviews among older adults to assess comprehension and cultural appropriateness. Phase 3 comprised pilot psychometric testing with a sample of older adults, excluding the Phase 2 participants.

Ethical Considerations

All study procedures, including recruitment, informed consent, data collection, and data management, were conducted in full accordance with relevant guidelines and regulations. The study protocol has been reviewed and approved by the University of the Philippines Manila Research Ethics Board (Protocol No.

UPMREB 2024-0781-01). The informed consent process was followed and signed consent forms were collected before proceeding with the data collection.

Study Setting and Participants

The study was conducted in Barangay UP Campus, Diliman, Quezon City, Philippines, an urban community with a population of approximately 8,500 residents, of whom approximately 800 are aged 60 years and older (De la Vega *et al.* 2021). This local community was chosen as the study site due to the existing partnership with the local government, ongoing community-based research program, and the active participation of the community residents and officials. Data collection was conducted between August and September 2025.

Phase 1 – Translation and Expert Review (n=11 stakeholders): Stakeholders included two bilingual translators for forward translation (CKC, GFR), two independent translators for backward translation (AOB, JDA), and a seven-member expert committee comprising one geriatric specialist, one gerontologist, one linguist, two public health specialists, one expert in Philippine studies, and one gerontologist who is also an older adult.

Phase 2 – Cognitive Interviews (n=12): Twelve older adults were recruited through purposive sampling to ensure diversity in age (60–85 years), sex, educational attainment (elementary to college level), and functional status (independent to mild mobility limitation). Participants were required to be Filipino-speaking, aged 60 years or older, and able to provide informed consent. Individuals with severe cognitive impairment, acute illness, or inability to communicate in Filipino were excluded.

Phase 3 – Pilot Psychometric Testing (n=80): Eighty older adults were recruited through community health worker referrals and senior citizen association announcements. Inclusion criteria were: age ≥60 years, Filipino-speaking, resident of Barangay UP Campus, able to provide informed consent, and not a participant of the ongoing and more extensive Qigong study. Exclusion criteria were: severe cognitive impairment, severe sensory impairment preventing participation in interviews, and inability to communicate in Filipino. A convenience subsample of ten participants was re-administered the questionnaire after a two-week interval to assess test-retest reliability. The test–retest subsample was exploratory and limited in size. The two-week interval is commonly utilized in studies to assess test-retest reliability to prevent memory recall of answers while minimizing changes in the constructs being measured (Briere 2011; Chiu *et al.* 2019).

Translation Procedure (Phase 1)

The translation process followed the guidelines of Beaton *et al.* (2000) and Guillemin *et al. (*1995). Two independent bilingual translators (CJC & GFR), with Filipino as their first language, completed forward translations from English to Filipino. The two forward translations were compared and reconciled through discussion to produce a single preliminary Filipino version. Another set of two independent bilingual translators (AOB & JDA), both English speakers and fluent in Filipino, completed backward

translations from Filipino to English without access to the original English questionnaire. The research team compared the backward translations with the original English version to identify discrepancies in meaning. A total of twelve semantic discrepancies were identified and discussed.

The forward translations, backward translations, and original English version were then reviewed by a seven-member expert committee. The committee assessed semantic equivalence (the literal meaning of words), idiomatic equivalence (colloquial expressions), experiential equivalence (the relevance of items to daily experiences), and conceptual equivalence (construct validity within the Filipino cultural context) (Guillemin *et al.* 1995; Beaton *et al.* 2000). Detailed translation procedures can be found in Appendix I.

Content Validity Assessment

Each of the seven expert committee members independently rated each item for relevance to its intended TFA construct using a 4-point scale (1 = not relevant, 2 = somewhat relevant, 3 = quite relevant, 4 = highly relevant) (Oliveira *et al.* 2026). Details on validation procedures can also be found in Appendix I. The item-level content validity index (I-CVI) was computed as the proportion of experts who rated the item as 3 or 4. Scale-level content validity index based on the average method (S-CVI/Ave) was computed as the mean of all I-CVIs. Scale-level content validity index based on universal agreement (S-CVI/UA) was computed as the proportion of items rated 3 or 4 by all experts.

Cognitive Interviews (Phase 2)

Cognitive interviews were conducted with twelve older adults using a think-aloud technique (Künzli 2009). Participants were asked to read each item aloud, verbalize their thought processes as they formulated a response, and provide their rating. After completing all items, the interviewer (ZGR) probed for any items that were confusing, difficult to answer, or contained unfamiliar words. Interview notes were coded by two independent researchers (APG and ZGR). Discrepancies were resolved through discussion. Consistent with the recommended review and finalization procedures for cognitive debriefing, items repeatedly identified as problematic were revised by consensus of the committee. Because no universally accepted cutoff exists for revising items based on cognitive debriefing feedback, we prespecified a pragmatic rule that items flagged by at least 3 of 12 participants would undergo committee review and revision (Wild *et al.* 2005). Necessary revisions were made as described above.

Pilot Testing and Instrumentation (Phase 3)

The final Filipino TFA questionnaire consisted of ten items, with one item per TFA construct except for Affective Attitude and Ethicality consisting of two items (Q1 & Q2=Affective Attitude, Q3=Burden, Q4 & Q5=Ethicality, Q6=Perceived Effectiveness [overall benefit], Q7=Intervention Coherence [understanding], Q8=Self-Efficacy, Q9=Opportunity Costs) plus one item assessing overall acceptability (Q10=General Acceptability). The items were rated on a 5-point Likert scale (1= Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly Agree). Higher scores indicate greater acceptability. Items were adapted to assess an example intervention, the acceptability of Qigong meditation, a gentle mind-body exercise

program described to participants as involving slow movements, breathing exercises, and meditation (Stahl *et al.* 2020; Yeh *et al.* 2025).

Questionnaires were administered face-to-face by two research assistants trained by the investigators. Sociodemographic data and functional status were collected via structured interview. The questionnaires were administered through a traditional paper-based method. The responses were then encoded in an online form via Qualtrics. The investigators reviewed, processed, and cleaned the encoded responses. The data was endorsed to the statistician for analysis.

Statistical Analyses

Descriptive statistics (means, standard deviations, frequencies, percentages) were computed for sociodemographic characteristics and item responses. Floor and ceiling effects were assessed by computing the percentage of participants responding at the lowest or highest possible score, with >15% considered substantial (Gulledge *et al.* 2019). For content validity, I-CVI, S-CVI/Ave, and S-CVI/UA were computed as described above. For construct validity, exploratory factor analysis (EFA) with principal components extraction and varimax (orthogonal) rotation was conducted. For internal consistency, Cronbach's alpha was computed for the total scale, with $\alpha \geq 0.70$ considered acceptable (Adamson and Prion 2013). For test-retest reliability, Intraclass correlation coefficient (ICC) using a two-way mixed-effects model with absolute agreement was computed for the total scale score and individual items (McGraw and Wong 1996). Detailed statistical analysis, cutoff values, and interpretation can be found in Appendix II. All analyses were conducted using STATA version 19 (StataCorp, College Station, TX, USA). Statistical significance was set at $p<0.05$.

Results

Translation and Cultural Adaptation

Twelve semantic discrepancies were identified from the forward-backward translation process that required resolution by the expert committee. The expert committee engaged in extensive deliberation on five items: Q3 (Burden), Q4 and Q5 (Ethicality), Q8 (Self-Efficacy), and Q9 (Opportunity Costs). Key decisions included: (1) replacing *"interbensyon"* (intervention, borrowed from English and perceived as technical jargon) with "*ehersisyo o programa*" (exercise or program); (2) contextualizing *"moral/etikal"* (moral/ethical) with concrete examples (e.g., *"walang masamang epekto sa iba,"* meaning "no harmful effects on others") to ground the abstract concept; (3) replacing *"makatarungan"* (fair/just, formal/legalistic register) with "patas" (fair/even-handed, conversational and accessible); (4) replacing *"tiwala sa sarili"* (self-trust, which participants interpreted as trust in one's character) with *"kumpiyansa sa sarili"* (self-confidence, which participants correctly interpreted as belief in one's ability); and (5) adopting neutral framing for opportunity cost items to avoid triggering defensive or socially desirable responses (e.g., "significantly interferes" rather than "prevents you from doing other important things").

Content Validity

Table 1 presents the content validity indices for each item. Most items achieved perfect I-CVI scores of 1.00, indicating unanimous expert agreement that the item was highly relevant to its intended construct. Two items, Q3 (Burden) and Q5 (Ethicality), received lower universal agreement (UA) scores of 0.43 and 0.29, respectively, due to one or two experts rating these items as "quite relevant" (3) rather than "highly relevant" (4). However, both items achieved I-CVI ≥ 0.86 (6 of 7 experts rated 3 or 4), meeting the threshold for acceptable relevance. After iterative revision and re-rating, the final questionnaire achieved excellent scale-level content validity: S-CVI/Ave = 0.97 and S-CVI/UA = 0.80, both exceeding recommended thresholds**.**

Table 1. Content Validity Indices for the Filipino TFA Questionnaire (n=7 Expert Members).

Item	TFA Construct	I-CVI	UA
Q1	Affective Attitude	1.00	1.00
Q2	Affective Attitude	1.00	1.00
Q3	Burden	0.86	0.43
Q4	Ethicality	1.00	1.00
Q5	Ethicality	0.86	0.29
Q6	Perceived Effectiveness (general)	1.00	1.00
Q7	Intervention Coherence (understanding)	1.00	1.00
Q8	Self-Efficacy	1.00	1.00
Q9	Opportunity Costs	1.00	1.00
Q10	General Acceptability	1.00	1.00
Scale-level indices			
S-CVI/Ave		**0.97**	
S-CVI/UA			**0.80**

Note. I-CVI = Item-level Content Validity Index (proportion of experts rating item 3 or 4 on 4-point relevance scale); UA = Universal Agreement (proportion of experts rating item 4 "highly relevant"); S-CVI/Ave = Scale-level CVI based on average method; S-CVI/UA = Scale-level CVI based on universal agreement method. Thresholds: I-CVI ≥0.78, S-CVI/Ave ≥0.90, S-CVI/UA ≥0.80.

Cognitive Interviews

Twelve older adults (8 female, 4 male; mean age 67.5 years, range 60–82; education: 3 elementary, 5 high school, 4 college) participated in cognitive interviews. The think-aloud process identified comprehension difficulties with specific terminology. The word *"interbensyon"* was flagged by 10 of 12

participants (83%) as unfamiliar or overly technical; participants spontaneously suggested, *"ehersisyo"* (exercise), *"programa"* (program), or *"gawain"* (activity) as more accessible alternatives. Seven participants (58%) questioned the meaning of "moral o etikal" (moral or ethical), with one participant describing the term as *"malalim"* (deep/profound), indicating abstract complexity. Five participants (42%) noted that *"makatarungan"* sounded formal and "parang legal" (like legal language), preferring "patas" (fair). Six participants (50%) interpreted "tiwala sa sarili" as trust in one's moral character rather than confidence in one's ability to perform a behavior, leading them to adopt *"kumpiyansa sa sarili."* Three participants (25%) initially responded defensively to the opportunity cost item when it was phrased as "prevents you from doing important things," prompting neutral rephrasing to "significantly interferes with other activities." Considering the feedback from the interviews, the questionnaire items were modified. The final translated TFA questionnaire used for the pilot testing can be found in Appendix III.

Sociodemographic Characteristics of Pilot Sample

The eighty older adult samples who participated in the pilot psychometric study had a mean age of 69.3 years (SD=6.2, range 60–87), was predominantly female (75%), and had varied educational attainment, with the largest group having completed high school (46.3%). Most participants were married (47.5%) or widowed (35.0%). Primary income sources were pensions (43.8%), small businesses (27.5%), and family support (28.8%). Functional limitations were common: 36.3% reported mobility difficulties, 27.5% reported vision problems, and 27.5% reported hearing difficulties; 40.0% reported no functional limitations. Detailed sociodemographic and functional characteristics can be found in Appendix IV.a.

Construct Validity: Exploratory Factor Analysis

Preliminary tests confirmed the suitability of the data for factor analysis. The Kaiser Meyer Olkin measure was 0.76, indicating adequate sampling adequacy, while Bartlett's test of sphericity was significant, $\chi^2(45) = 184.26$, $p < 0.001$. These results supported the appropriateness of proceeding with factor analysis. The scree plot and eigenvalues supported a two-factor solution, with eigenvalues of 3.38 and 1.46. Together, these two factors explained 48.43% of the total variance. After varimax rotation, items loaded onto two components, as shown in Table 2.

Table 2. Exploratory factor analysis of the 10 item TFA (Filipino).

Item	TFA Construct	Factor 1: Perceived Value & Ethical Appraisal	Factor 2: Practical Engagement & Comprehension	h^2
Q1	Affective Attitude (Liking)	0.38		0.16
Q2	Affective Attitude (Comfort)	0.78		0.61
Q3	Burden		0.77	0.59
Q4	Ethicality	0.54	0.33	0.40
Q5	Ethicality (Fairness)		0.73	0.61
Q6	Perceived Effectiveness	0.81		0.67
Q7	Intervention Coherence		0.69	0.48
Q8	Self-Efficacy	0.42	0.49	0.41
Q9	Opportunity Costs	0.51		0.31
Q10	General Acceptability	0.63	0.46	0.60
Eigenvalue		**3.38**	**1.46**	
% Variance Explained		**33.78%**	**14.65%**	
Cumulative % Variance		**33.78%**	**48.43%**	

Note. *Extraction method is principal component analysis. Rotation is varimax method. Loadings < 0.30 were not displayed.*

Factor 1, labeled Perceived Value and Ethical Appraisal, consisted of items related to emotional responses, moral considerations, and overall judgments about the intervention's acceptability. This factor may be interpreted as reflecting participants' appraisal of the intervention's value and acceptability. Factor 2, labeled Practical Engagement and Comprehension, included items related to perceived effort, fairness, and clarity of the intervention. This factor may be interpreted as reflecting the practical and experiential aspects of acceptability.

Several items demonstrated moderate cross-loadings, particularly Q4, Q8, and Q10, although most communalities exceeded 0.40. Communalities ranged from 0.16 for Q1 to 0.67 for Q6, suggesting that most items were reasonably well represented by the two-factor model. This structure supports the

multidimensional nature of intervention acceptability, with the translated tool capturing both value-based and practical aspects of older adults' appraisal of the intervention.

The overall Cronbach's alpha for the 10-item scale was 0.73 (Appendix IV.b). Item-total correlations ranged from 0.16 (Q9, Opportunity Costs) to 0.63 (Q10, General Acceptability), with most items exceeding the recommended threshold of $r \geq 0.30$. Q10 (General Acceptability) and Q6 (Perceived Effectiveness) had the strongest item-total correlations ($r = 0.63$ and $r = 0.53$, respectively), consistent with their high loadings on Factor 1. Q9 (Opportunity Costs) had the lowest item-total correlation ($r = 0.16$), and its deletion would increase alpha to 0.77, suggesting it contributes less to overall scale coherence; however, given its theoretical importance within the TFA framework and the acceptable overall alpha, Q9 was retained.

Cronbach's alpha when an item was deleted ranged from 0.68 to 0.77, indicating that no single item substantially diminishes the scale's reliability (Appendix IV.b). The total mean scale score was 42.18 (SD = 3.73), reflecting high overall acceptability of the identified intervention among participants.

Item means ranged from 3.94 (Q9, Opportunity Costs) to 4.38 (Q7, Intervention Coherence), with all items exceeding the midpoint of 3.0, indicating generally high acceptability. Ceiling effects (proportion of respondents selecting "5 = Strongly Agree") were substantial for several items: Q7 (47.5%), Q6 (40.0%), Q8 (37.5%), Q3 (36.2%), Q10 (33.8%), Q5 (31.2%), Q4 (28.8%), Q9 (28.8%), Q1 (27.5%), and Q2 (23.8%). These high ceiling effects reflect the very positive reception of Qigong in this sample but also raise concerns about the instrument's discriminative capacity in populations with universally favorable attitudes toward a given intervention. Floor effects were minimal (<2%) for all items, indicating no response restriction at the lower end of the scale. Full item statistics can be found in Appendix IV.b.

Inter-Item Correlations

Spearman's inter-item correlation matrix provided additional evidence of the scale's internal structure (Table 3). Correlations ranged from −0.02 (Q1 and Q9) to 0.49 (Q3 and Q5; Q3 and Q7), indicating a generally low to moderate positive correlation pattern appropriate for a multi-construct instrument. The strongest associations were observed between Q3 (Burden) and Q5 (Ethicality, fairness) ($\rho = 0.49$), and between Q3 (Burden) and Q7 (Intervention Coherence) ($\rho = 0.49$), consistent with their clustering under Factor 2. In contrast, Q9 (Opportunity Costs) showed the weakest correlations with the other items, ranging from −0.02 to 0.29, further supporting its relative independence within this sample. No item pair exceeded $\rho = 0.85$, indicating the absence of problematic redundancy and supporting the retention of all items as contributing distinct information.

Table 3. Spearman's Inter-Item Correlation Matrix (n=80)

Item	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
Q1	**(0.73)**									
Q2	0.21	**(0.70)**								
Q3	0.17	0.08	**(0.72)**							
Q4	0.12	0.36	0.31	**(0.70)**						
Q5	0.04	0.28	0.49	0.39	**(0.69)**					
Q6	0.31	0.43	0.16	0.43	0.32	**(0.69)**				
Q7	0.22	0.22	0.49	0.17	0.41	0.26	**(0.72)**			
Q8	0.13	0.24	0.31	0.31	0.44	0.26	0.25	**(0.70)**		
Q9	-0.02	0.21	0.02	0.29	0.05	0.22	0.05	0.14	**(0.77)**	
Q10	0.29	0.47	0.41	0.43	0.48	0.28	0.28	0.36	0.18	**(0.68)**

Note. Off-diagonal values represent inter-item correlations.

Diagonal values in parentheses are the Cronbach's alpha of the scale when the corresponding item is removed.

Test-Retest Reliability

Item-level ICCs from a subset of 10 participants varied widely, ranging from –0.62 (Q8, self-efficacy) to 0.38 (Q4, ethicality), indicating that most items showed poor to moderate consistency across administrations. Items Q2 (comfort, ICC = 0.37) and Q4 (ethicality, ICC = 0.38) showed the strongest stability, whereas Q3 (burden, ICC = –0.34) and Q8 (self-efficacy, ICC = –0.62) indicated instability over time. Several items, such as Q5 (fairness) and Q6 (perceived effectiveness), produced near-zero ICCs, suggesting little agreement between baseline and follow-up responses (Appendix IV.c).

Discussion

The forward-backward translation process identified twelve semantic discrepancies that required resolution by the expert committee. The most contentious terms were those reflecting abstract psychological constructs, consistent with challenges reported in other TFA adaptations (Judan-Ruiz *et al.* 2020; Torres and Macindo 2020; Rivera *et al.* 2025). The content validity of the Filipino TFA questionnaire substantially exceeded that reported for other cross-cultural adaptations such as the Spanish and Swedish versions (Rivera *et al.* 2025; Samielsson *et al.* 2025). Comparable Filipino-language validation studies have similarly demonstrated that rigorous multi-phase adaptation processes yield high content validity indices: the Filipino Lower Extremity Functional Scale achieved S-CVI/Ave > 0.90 (Nuñes *et al.* 2022), the Filipino Harris Hip Score achieved S-CVI of 0.963 (Tan *et al.* 2025), Spanish version

reported McDonald's omega = 0.60 (Rivera *et al.* 2025), and the Filipino Vertigo Handicap Questionnaire achieved S-CVI of 0.99 (Orpilla *et al.* 2023), collectively affirming that systematic expert committee review and iterative revision are essential to achieving excellent content validity in Filipino-language instruments.

The superior content validity of the Filipino TFA version likely reflects several methodological strengths: the use of a multidisciplinary expert committee (n=7); the iterative revision process whereby Items 3 (Burden, UA=0.43) and 5 (Ethicality, UA=0.29) that initially showed low universal agreement were substantially revised and re-rated; and the extensive cognitive interview phase (n=12) that identified specific problematic terms and guided linguistically and culturally appropriate adaptations. This multi-phase approach, incorporating forward-backward translation, expert committee review, cognitive interviews, and pilot testing, is consistent with internationally recognized guidelines for cross-cultural adaptation of health measures (Veloso and Arandia 2024; Manalo *et al.* 2025). The Swedish TFA pilot (Samuelsson *et al.* 2025) similarly employed cognitive interviews using verbal probing and think-aloud techniques and the Danish TFA adaptation (Bekker-Jensen *et al.* 2026) likewise identified four main issues through cognitive interviews.

Items reflecting ethicality and self-efficacy were the most challenging to translate and validate across cultural contexts. The Swedish version struggled with the wording of the "comfortable" items in affective attitude (Samuelsson *et al.* 2025), paralleling our findings that "moral/etikal" required contextualization and that opportunity cost items required neutral framing to avoid triggering defensive responses. These cross-cultural consistencies suggest that certain TFA constructs, particularly ethicality, self-efficacy, and opportunity costs, are inherently abstract and value-laden, requiring careful linguistic and cultural adaptation for each target population (Gomez *et al.* 2021; Veloso and Arandia 2024).

Cultural and linguistic factors played a central role in shaping translation decisions and may explain some of the unique challenges encountered in the Filipino context (Babayev 2023). The Filipino language contains multiple registers ranging from formal/academic to conversational/colloquial (Roxas *et al.* 2008), and older adults with modest educational backgrounds typically prefer the conversational register. The initial use of formal terms such as "makatarungan" (justice/fairness) was perceived as overly legalistic and inaccessible; the substitution of "patas" (fair/even-handed) resolved this issue. Similarly, "interbensyon" (a direct borrowing from English "intervention") was unfamiliar to 83% of cognitive interview participants, whereas "ehersisyo" (exercise) and "programa" (program) were immediately understood.

These findings are consistent with experiences reported in other Filipino-language adaptation studies. Garabiles *et al.* (2019); similarly found that pure Filipino translations of abstract psychological terms "sounded odd or were incomprehensible" to Filipino participants, leading to the adoption of "Taglish" (merged English and Filipino) when necessary. The Sensory Processing and Self-Regulation Checklist–Tagalog adaptation (Gomez *et al.* 2021) likewise incorporated colloquially adapted English terms due to the lack of equivalent Filipino terms or because direct translation led to loss of meaning. These findings

underscore the importance of selecting vocabulary that matches the target population's linguistic experience, not merely achieving literal semantic equivalence (Garabiles *et al.* 2019).

Cultural adaptation of health interventions for Filipino populations must also account for the role of language in shaping acceptability and comprehension. Studies have demonstrated that using the native language and culturally familiar metaphors is essential for delivering acceptable and appropriate mental health literacy programs to Filipino populations, with evidence suggesting that using participants' local language is an effective way to achieve positive intervention outcomes (Martinez *et al.* 2022; Wang *et al.* 2024). Similarly, other studies found that culturally and linguistically adapted mobile health interventions demonstrated high feasibility and acceptability among Tagalog-speaking Filipino Americans, underscoring the importance of linguistic adaptation beyond mere translation (Martinez *et al.* 2022; Nguyen and Lavretsky 2020).

The two-factor structure that emerged from exploratory factor analysis diverged from the theoretical seven-construct TFA framework. Factor 1 ("Perceived Value & Ethical-Appraisal") was defined by items reflecting perceived effectiveness, ethicality, self-efficacy, and general acceptability, suggesting that Filipino older adults may integrate judgments about an intervention's likely benefits, moral appropriateness, and their own capability into a holistic evaluation of its overall value. Factor 2 ("Practical Engagement & Comprehension") was defined by items reflecting burden and intervention coherence, suggesting that the effort required and the understandability of an intervention constitute a distinct dimension of acceptability focused on the practicalities of participation. This factor structure is not entirely unexpected, as the Swedish TFA validation (Samuelsson *et al.* 2025) similarly reported that 7 out of 10 items were below threshold values for content validity, indicating that the seven-construct model does not always map cleanly onto empirical factor structures across cultural contexts. Moreover, research on multidimensional constructs suggests that individuals with lower formal education or less familiarity with abstract psychological constructs tend to respond more holistically, collapsing distinctions between closely related constructs (Johnson *et al.* 2012; Polites *et al.* 2012). Given that 46.3% of our sample had a high school education or less, it is plausible that participants did not sharply differentiate between, for example, affective attitude and perceived effectiveness, but instead integrated these into a general "this seems good" judgment.

The internal consistency reliability of the Filipino TFA questionnaire (Cronbach's $\alpha = 0.73$) was within the acceptable range (Nunnally 1978) and comparable to that reported in other Filipino-language validation studies (Veloso and Arandia 2024; Manalo *et al.* 2025; Tan *et al.* 2025), collectively demonstrating that α values in the range of 0.70–0.75 are typical and appropriate for brief, multidimensional Filipino-language health questionnaires. For a brief (10-item) multidimensional scale assessing theoretically distinct constructs, $\alpha = 0.70–0.80$ is appropriate; values exceeding 0.90 would suggest item redundancy and raise concerns that the scale is unidimensional when it should be multidimensional (Johnson *et al.* 2012; Torres and Macindo 2020; Tan *et al.* 2025).

The test-retest reliability of the Filipino TFA questionnaire was poor (ICC = −0.06, n=10), but this finding cannot be considered definitive evidence of temporal instability given the critically small sample size. A minimum of 30 participants is recommended for ICC estimation to achieve adequate power and precision (Bujang and Baharum 2017); our pilot study included only 10 participants for the retest subsample due to resource and time constraints. The 95% CIs for all ICCs were extremely wide, ranging from negative to moderate positive values, indicating that the estimates are statistically unreliable. The wide confidence intervals around the item-level estimates likely reflect the small retest sample size (n = 10), limiting the precision and interpretability of these findings. Nonetheless, the results provide preliminary insight into the reliability of the Filipino TFA questionnaire, indicating the need for further testing with a larger sample to accurately assess its temporal stability.

Other Filipino-language validation studies have similarly encountered challenges in estimating test-retest reliability in small samples (Chiong Maya *et al.* 2023; Herras *et al.* 2022), suggesting that adequate sample size is critical for reliable ICC estimation. The Swedish TFA pilot (Samuelsson *et al.* 2025); similarly acknowledged limited test-retest data and recommended larger samples for future validation. We strongly recommend that future studies assess test-retest reliability with sample sizes of n≥50, standardized retest intervals (7–14 days), and screening to exclude participants who experience major health or life events between assessments.

Multiple items showed ceiling effects, with Q7 (Intervention Coherence-understanding) exhibiting the most pronounced. Ceiling effects indicate that a substantial proportion of the sample rated the intervention at the upper limit of the scale, thereby reducing the instrument's ability to detect differences between groups or changes over time. The high ceiling effects observed in this study likely reflect genuinely high acceptability of the Qigong intervention in this sample, as Qigong is a gentle, culturally familiar practice that requires minimal equipment and has documented benefits for older adults.

Cultural adaptation studies of health interventions have consistently found that interventions perceived as culturally relevant and socially valued tend to generate high acceptability ratings with limited variability (Garabiles *et al.* 2019; Chiong Maya *et al.* 2023). One study similarly noted that cultural adaptation of obesity-related behavioral prevention interventions increases acceptability among target cultural groups, though this can reduce the discriminative sensitivity of acceptability measures (Bekker-Jensen *et al.* 2026). Researchers planning to use the Filipino TFA questionnaire should be aware that in samples with very positive attitudes toward an intervention, the scale may show limited variability and reduced sensitivity to detect subgroup differences or predictors of acceptability.

The implications of this research are significant for health services research and public health policy in the Philippines. The Filipino TFA questionnaire provides a standardized, theoretically informed, and culturally appropriate tool that can be administered in approximately five minutes, making it feasible for inclusion in program evaluations, clinical trials, and implementation science studies. It can be integrated into routine monitoring by local government unit (LGU) health offices, community health centers, senior citizen associations, and nongovernmental organizations delivering health services to older Filipinos.

Future research priorities include: (1) conducting confirmatory factor analysis with a larger independent sample (n≥200) to test the stability and fit of the two-factor structure; (2) assessing test-retest reliability with an adequately powered sample (n≥50) under standardized conditions (3) conducting multi-site and muti-language validation across urban and rural settings and major linguistic regions (Tagalog, Cebuano, Ilocano) to establish broader generalizability; (4) applying the questionnaire to diverse intervention types to assess its generic applicability (Arboleda 2026); (5) conducting measurement invariance testing across key subgroups; and (6) conducting longitudinal studies to examine whether baseline TFA scores prospectively predict intervention uptake, adherence, and retention, thereby establishing predictive validity and demonstrating the clinical utility of the measure.

Conclusions

The Filipino version of the Theoretical Framework of Acceptability (TFA) questionnaire is the first validated Filipino-language instrument for assessing the acceptability of healthcare interventions among older adults. Demonstrating excellent content validity (S-CVI/Ave = 0.97) and acceptable internal consistency (Cronbach's α = 0.73), it provides a brief, theoretically informed, and culturally appropriate tool for Philippine health researchers, clinicians, and policymakers. The two-factor structure: "Perceived Value & Ethical-Appraisal" and "Practical Engagement & Comprehension", reflects the cognitive and cultural dimensions through which Filipino older adults evaluate intervention acceptability. While test-retest reliability requires further validation with larger samples, the instrument is ready for use in program evaluation, implementation science research, and health policy applications. Its availability fills a critical gap in the aging and health research infrastructure of an aging society, supporting evidence-based efforts to enhance the acceptability and effectiveness of health interventions for Filipino older adults.

Declarations

Acknowledgements

The authors thank the National Institutes of Health, University of the Philippines Manila, for funding the study (Grant No. NIH 2025-009), the Barangay UP Campus officials and senior citizen association officers for facilitating community access and participant recruitment, and the UP Manila units for providing administrative assistance to the project. We are grateful to the expert committee members and all study participants for their time and valuable contributions. We also thank the authors who developed the original TFA framework and generic questionnaire, which served as the foundation for this work.

References

1. ADAMSON KA. PRION S. 2013. Reliability: measuring internal consistency using Cronbach's α. Clinical Simulation in Nursing 9(5): e179–e180.

2. ARBOLEDA J. 2026. Linguistic Validation and Cross-Cultural Adaptation of the Shoulder Telehealth Assessment Tool for Filipino Patients with Musculoskeletal Shoulder Condition: Cross-Sectional Study. JMIR Rehabilitation and Assistive Technologies 13: e67974–e67974
3. BABAYEV J. 2023. Impact of socio-linguistic and socio-cultural factors on translation process. Sciences of Europe (128):50–53.
4. BEATON DE. *et al*. 2000. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine 25(24): p. 3186–3191.
5. BEKKER-JENSEN D. *et al.* 2026. Danish cross-cultural adaptation of the Theoretical Framework of Acceptability questionnaire. Danish Medical Journal 73(3).
6. BELARGA CA *et al.* 2022. Noncommunicable Disease (NCD) Treatment, prevention strategies, and risk factors: A Philippine situation. International Journal of Research Publication and Reviews 2582:7421.
7. BRIERE JL. 2011. Determining the Psychometric Properties of the Retrieval-Induced Forgetting Procedure. 2011, University of Saskatchewan.
8. BUJANG MA. BAHARUM N. 2017. A simplified guide to determination of sample size requirements for estimating the value of intraclass correlation coefficient: a review. Archives of Orofacial Sciences 12(1).
9. CASTILLO AAP. 2024. Promotion and control of chronic non-communicable diseases in older adults: A Bibliographical Review. Centro Sur 8(2).
10. CHIONG MAYA AC. *et al.* 2023. Translation and Cross-Cultural Adaptation of English Version of Kiddy-KINDL to Filipino Language in Assessing the Children's Health-Related Quality of Life (CHRQoL). International Journal of Allied Health Sciences and Practice 21(3):18.
11. CHIU EC. *et al.* 2019. Test-retest reliability and minimal detectable change of the Cognitive Abilities Screening Instrument in patients with dementia. PLoS One 14(5): e0216450.
12. DE LA VEGA SAF *et al.* 2021. Active aging health determinants among working and retired Filipino older persons living in an urban academic campus. Act Med Philip 55(4).
13. GARABILES MR. SHEHADEH MH. HALL BJ. 2019 Cultural Adaptation of a Scalable World Health Organization E-Mental Health Program for Overseas Filipino Workers. JMIR Formative Research 3(1): e11600.
14. GOMEZ IN. MORATO-ESPINO PGG. LAI CYY. 2021. Examining the Linguistic Equivalency and Cross-Cultural Adaptation of the Sensory Processing and Self-Regulation Checklist- Tagalog Version. Asian Journal of Occupational Therapy 17(1):57–63.
15. GOMEZ INB. *et al.* 2021. Psychometric Properties of the Sensory Processing and Self-Regulation Checklist: English Version. Occupational therapy international (1): 6658786.
16. GRININ L. GRININ A. KOROTAYEV A. 2023. Global aging: An integral problem of the future. How to turn a problem into a development driver? In Reconsidering the limits to growth: A report to the Russian association of the club of Rome. Springer. 117–135.

17. GUILLEMIN F. *et al.* 1995. Cross-cultural adaptation and validation of health status measures. Scandinavian Journal of Rheumatology 24(2): 61–63.
18. GULLEDGE CM. *et al.* 2019. Floor and ceiling effects, time to completion, and question burden of PROMIS CAT domains among shoulder and knee patients undergoing nonoperative and operative treatment. Journal of Bone and Joint Surgery Open Access 4(4): e0015.
19. HE W. *et al.* 2022. Asia aging: demographic, economic, and health transitions. US Department of Commerce, Washington, DC, USA, Report 1.
20. HERRAS ARR. *et al.* 2022.Cross-Cultural Adaptation and Reliability of the Multidimensional Outcome Expectations for Exercise Scale (MOEES) in Filipino Older Adults. Philippine Journal of Physical Therapy 1(1).
21. IZAH SC. SYLVA L. HAIT M. 2023. Cronbach's alpha: A cornerstone in ensuring reliability and validity in environmental health assessment. Engineered Science Energy and Environment 23:1057.
22. JOHNSON RE. *et al.* 2012. Recommendations for improving the construct clarity of higher-order multidimensional constructs. Human Resource Management Review 22(2): 62–72.
23. JUDAN-RUIZ EA. MINA RJL. MACINDO JRB. 2020. Psychometric Properties of the Filipino Version of Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS): A Cross-Cultural Validation Study. Journal of Patient Experience 7(6): 1526–1534.
24. KOO TK. LI MY. 2016. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. Journal of chiropractic medicine 15(2): p. 155–163.
25. KÜNZLI A. 2009. Think-aloud Protocols–A Useful Tool for Investigating the Linguistic Aspect of Translation 1. Meta 54(2): 326–341.
26. MANALO APS. *et al.* 2025. Filipino translation and cross-cultural adaptation of the Diabetic Foot Knowledge Subscale (DFKS) and Foot Self-Care Behavior Scale (FSCBS) and its content validation and reliability testing. Acta Medica Philippina 59(16): 7.
27. MARTINEZ AB. *et al.* 2022. 'C'mon, Let's Talk: A Pilot Study of Mental Health Literacy Program for Filipino Migrant Domestic Workers in the United Kingdom. Social Psychiatry and Psychiatric Epidemiology 59(2): 385–401.
28. MCGRAW KO. WONG SP. 1996. Forming inferences about some intraclass correlation coefficients. Psychological Methods 1(1): 30.
29. MOKKINK LB. *et al.* 2016. The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) and how to select an outcome measurement instrument. Brazilian Journal of Physical Therapy 20(2):105–113.
30. MOOI E. SARSTEDT M. MOOI-RECI I. 2017. Principal component and factor analysis, in Market research: The process, data, and methods using stata. Springer. 265–311.
31. NGUYEN S. LAVRETSKY H. 2020. Emerging Complementary and Integrative Therapies for Geriatric Mental Health. Current treatment options in psychiatry 7(4): 447–470.

32. NUÑEZ AS. *et al.* 2022. Filipino Translation and Cross-Cultural Adaptation of the Lower Extremity Functional Scale. Philippine Journal of Physical Therapy 1(1).
33. NUNNALLY JC. 1978. An Overview of Psychological Measurement. In: Wolman, B.B. (eds) Clinical Diagnosis of Mental Disorders. Springer, Boston, MA.
34. OLIVEIRA AF. *et al.* 2026. Development and Preliminary Psychometric Evaluation of the ACRIS-CS: A Theory-Based Measure of Prospective Acceptability of Cognitive Rehabilitation in Cancer Survivors. Journal of clinical medicine, 15(5), 1858.
35. ORPILLA VB. *et al.* 2023. Translation and Cultural Adaptation of the Vertigo Handicap Questionnaire in Vestibular Rehabilitation of Older Filipino Adults. Philippine Journal of Physical Therapy 2(1): 46–46.
36. PATIL VH. *at al.* 2008. Efficient theory development and factor retention criteria: Abandon the 'eigenvalue greater than one' criterion. Journal of Business Research 61(2): 162–170.
37. POLIT DF. BECK CT. 2006. The content validity index: are you sure you know what's being reported? Critique and recommendations. Research in Nursing & Health 29(5): 489–497.
38. POLITES GL., ROBERTS N. THATCHER J. 2012. Conceptualizing models using multidimensional constructs: a review and guidelines for their use. European Journal of Information System 21(1): 22–48.
39. RIVERA S. *et al.* 2025. Adaptación transcultural y validación de un cuestionario genérico de aceptabilidad al español. Revista de Investigación e Innovación en Ciencias de la Salud 7(1): 1–19.
40. ROXAS REO. *et al.* 2008. Building language resources for a Multi-Engine English-Filipino machine translation system. Language Resources and Evaluation 42(2): 183–195.
41. RUPARELIYA DA. SHUKLA YU. 2020. Need for cross-cultural adaptation of self-reported health measures: review study. Indian Journal of Physiotherapy and Occupationa Therapy 14: 34–37.
42. SAMUELSSON M, MÖLLERBERG ML. NEZIRAJ M. 2025. The Swedish theoretical framework of acceptability questionnaire: translation, cultural adaptation, and descriptive pilot evaluation. BMC Health Service Research 25(1): 684.
43. SEKHON M. CARTWRIGHT M. FRANCIS JJ. 2017. Acceptability of healthcare interventions: an overview of reviews and development of a theoretical framework. BMC Health Service Research 17: 1–13.
44. SEKHON M. CARTWRIGHT M. FRANCIS JJ. 2022. Development of a theory-informed questionnaire to assess the acceptability of healthcare interventions. Health Service Research 22(1): 279.
45. SHI J. MO X. SUN Z. 2012. Content validity index in scale development. Journal of Central South University. Medical Sciences 37(2): 152–155.
46. SHRESTHA N. 2021. Factor analysis as a tool for survey analysis. American Journal of Applied Mathematics and Statistics 9(1): 4–11.
47. STAHL JE, BELISLE SS. ZHAO W. 2020. Medical Qigong for mobility and balance self-confidence in older adults. Frontiers in Medicien 7:422.

48. TAN DM. *et al.* 2025. Psychometric Properties of the Filipino Version of Harris-Hip Score Among Patients With Osteoarthritis and Hip Fractures. Philippine Journal of Orthopaedics 40(2).
49. TOBIAS S. CARLSON JE. 1969. Brief report: Bartlett's test of sphericity and chance findings in factor analysis. Multivariate Behavior Research 4(3): 375–377.
50. TORRES G. MACINDO JRB. 2020. Cross-Cultural Validation Into Filipino of the Consumer Assessment of Health-Care Providers and Systems Surgical Care Survey (S-Cahps). Journal of Patient Experience 7(6): 1136–1143.
51. VASQUEZ K. *et al.* 2001. Cultural variation and similarity in moral rhetorics: Voices from the Philippines and the United States. Journal of Cross-Cultural Psychology 32(1): 93–120.
52. VELOSO AL. ARANDIA E. 2024. Cross-Cultural Adaptation and Validation of the International Knee Documentation Committee Subjective Knee Form in Filipino. Philippine Journal of Orthopaedics 39(1):31–35.
53. WANG Y. *et al.* 2024. Exploring Unmet Healthcare Needs and Associated Inequalities Among Middle-Aged and Older Adults in Eastern China During the Progression Toward Universal Health Coverage. Health Economics Review 14(1).
54. WILD D. *et al.* 2005. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR task force for translation and cultural adaptation. Value in health, 2005. 8(2): 94–104.
55. YEH GY. *et al.* 2025. The Science of Tai Chi and Qigong as Whole Person Health-Part II: Evidence Gaps and Opportunities for Future Research and Implementation. Journal of Integrative and Complementary Medicine 31(6): 521–534.
56. ZHANG Z. *et al.* 2024. A mixed-methods data approach integrating importance-performance analysis (IPA) and 0-Meyer-Olkin (KMO) in applied talent cultivation. Journal of Applied Data Science 5(1): 256–267.
57. ZIJLMANS EA. *et al.* 2019. Item-score reliability as a selection tool in test construction. Frontiers in Psychology 9: 2298.

Figures

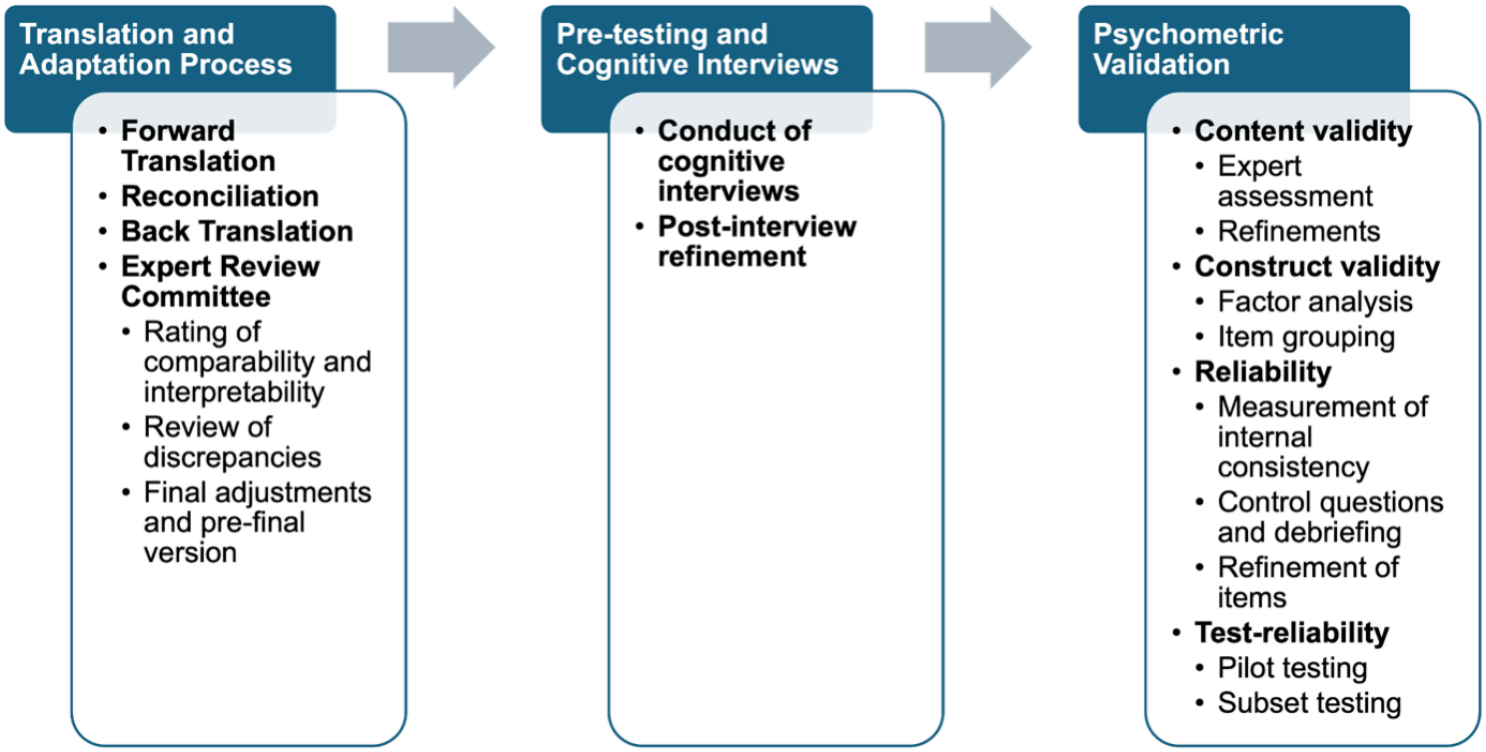


Figure 1

Legend not included with this version.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryfilesupdatedAppendicesTFApreprint.docx