

A Psychometric Evaluation Framework for Mobile Ux: Developing a Validated Questionnaire for Assessing Cognitive Load in Elderly Smartphone Users

¹Bhaskar Mani Tiwari, ²Prof. Swati V Chande

¹P.hD Scholar, Deptt. of Computer Science and IT, IIS University, Jaipur

²H.O.D, Deptt. of Computer Science and IT, IIIM, Jaipur

ARTICLE INFO

Received: 03 Aug 2026

Accepted: 21 Sept 2026

Published: 03 Oct 2026

ABSTRACT

The rapid proliferation of smartphone technology has created significant usability challenges for elderly users, particularly regarding cognitive load management during mobile interactions. This study presents the development and psychometric validation of the Elderly Smartphone Cognitive Load Questionnaire (ESCLQ), a novel instrument designed to assess cognitive workload in mobile user experience contexts for older adults aged 60 and above. Drawing upon established cognitive load theory, the NASA Task Load Index framework, and age-specific mobile usability principles, we developed a 24-item questionnaire across four dimensions: perceptual processing, motor-cognitive coordination, navigation complexity, and emotional-cognitive strain. The validation process involved 312 elderly participants (mean age = 68.4 years, SD = 6.2) across two phases: exploratory (n=156) and confirmatory (n=156). Exploratory factor analysis revealed a stable four-factor structure explaining 71.3% of the total variance, while confirmatory factor analysis demonstrated excellent model fit indices (CFI = 0.94, TLI = 0.93, RMSEA = 0.06). The ESCLQ demonstrated high internal consistency (Cronbach's α = 0.89) and strong test-retest reliability (ICC = 0.85) over a two-week interval. Convergent validity was established through significant correlations with the System Usability Scale (r = -0.67, p < 0.001) and NASA-TLX (r = 0.72, p < 0.001). Discriminant validity analyses confirmed the questionnaire's ability to differentiate between varying levels of interface complexity and user expertise. This validated instrument provides researchers and practitioners with a reliable, psychometrically sound tool for evaluating cognitive load in elderly smartphone users, facilitating the development of age-appropriate mobile interfaces and contributing to digital inclusion efforts for aging populations.

Keywords: Cognitive load assessment, elderly smartphone usability, psychometric validation, mobile user experience, age-friendly interface design

1. Introduction

1.1 Background and Motivation

The global demographic shift toward an aging population has profound implications for technology design and accessibility. By 2050, the proportion of individuals aged 60 and above is projected to reach 2.1 billion worldwide, representing approximately 22% of the global population (United Nations, 2019). Concurrently, smartphone penetration among older adults has increased dramatically, with recent studies indicating that over 61% of individuals aged 65 and above now own smartphones (Pew Research Center, 2021). However, this technological adoption is accompanied by significant usability challenges that disproportionately affect elderly users.

Age-related cognitive changes, including declines in working memory capacity, processing speed, and attentional control, fundamentally alter how older adults interact with mobile interfaces (Salthouse, 2019). Research indicates that cognitive impairment prevalence among non-smartphone users (17.8%) is

significantly higher than among smartphone users (1.4%), suggesting a critical relationship between mobile technology engagement and cognitive health. These findings underscore the importance of designing smartphone interfaces that minimize cognitive load while maximizing usability for elderly populations.

Despite growing recognition of these challenges, there exists a notable gap in validated assessment instruments specifically designed to measure cognitive load in elderly smartphone users. Existing tools such as the NASA Task Load Index (NASA-TLX) and System Usability Scale (SUS), while psychometrically robust, were not developed with the unique cognitive, perceptual, and motor characteristics of older adults in mind. This limitation hinders both research efforts to understand elderly mobile user experience and practical attempts to design age-appropriate interfaces.

1.2 Research Gap and Objectives

Current cognitive load assessment methods in human-computer interaction predominantly employ generic instruments that fail to capture age-specific interaction challenges. Studies examining smartphone usability for elderly users have identified critical barriers including small touch targets, complex navigation hierarchies, inadequate visual contrast, and insufficient error recovery mechanisms (Petrovčič et al., 2018; Gomez-Hernandez et al., 2023). However, quantitative assessment of the cognitive burden imposed by these design deficiencies remains limited due to the absence of validated, context-specific measurement tools.

Furthermore, while the NASA-TLX has demonstrated reliability in measuring cognitive workload in older adults (ICC = 0.73), its six-dimensional structure (mental demand, physical demand, temporal demand, effort, performance, frustration) does not adequately address the mobile-specific interaction patterns and age-related cognitive-motor integration challenges characteristic of elderly smartphone use. Similarly, the System Usability Scale, though validated across various digital contexts, lacks the granularity necessary to identify specific sources of cognitive load in mobile interfaces.

The primary objective of this research is to develop and validate a comprehensive psychometric instrument the Elderly Smartphone Cognitive Load Questionnaire (ESCLQ) specifically designed to assess cognitive workload in mobile user experience contexts for older adults. The specific aims include:

1. To develop a theoretically grounded questionnaire that captures the multidimensional nature of cognitive load in elderly smartphone interactions
2. To establish the factor structure and internal consistency of the instrument through exploratory and confirmatory factor analyses
3. To evaluate the reliability of the questionnaire using test-retest methods
4. To assess convergent and discriminant validity through correlations with established usability and cognitive load measures
5. To provide normative data and practical guidelines for researchers and UX practitioners working with elderly populations

1.3 Theoretical Framework

The development of the ESCLQ is grounded in three complementary theoretical perspectives: Cognitive Load Theory (CLT), the Technology Acceptance Model for seniors (STAM), and mobile usability heuristics for older adults.

Cognitive Load Theory, originally proposed by Sweller (1988), posits that learning and task performance are constrained by the limited capacity of working memory. CLT distinguishes between intrinsic load (inherent task complexity), extraneous load (imposed by instructional design), and germane load (devoted

to schema construction). In the context of elderly smartphone use, age-related declines in working memory capacity mean that even moderate extraneous load such as that imposed by poor interface design can significantly impair task performance and user experience (Chen et al., 2017).

The Senior Technology Acceptance Model (STAM), an extension of the original Technology Acceptance Model (TAM), incorporates age-specific factors including cognitive capabilities, physical conditions, social influences, and emotional needs as key predictors of perceived usefulness and ease of use (Chen and Chan, 2014). This framework recognizes that older adults' technology adoption is mediated by factors distinct from those affecting younger users, necessitating specialized assessment approaches.

Mobile usability heuristics for older adults emphasize the importance of simplified navigation, enhanced visual clarity, larger touch targets, error tolerance, and cognitive support features (Gomez-Hernandez et al., 2023; Petrovčič et al., 2018). These design principles inform the operational definition of cognitive load factors assessed by the ESCLQ, ensuring alignment between theoretical constructs and practical design considerations.

1.4 Significance and Contributions

This research makes several significant contributions to the fields of human-computer interaction, gerontechnology, and psychometric assessment. First, it addresses a critical gap in the methodological toolkit available to researchers studying elderly technology use by providing a validated, age-appropriate cognitive load assessment instrument. Second, it advances theoretical understanding of how cognitive load manifests in mobile contexts for older adults, potentially informing refined models of age-related technology acceptance and usability. Third, it offers practical value to UX designers and mobile application developers seeking evidence-based methods to evaluate and improve interface designs for elderly users.

The ESCLQ enables systematic assessment of cognitive workload across different mobile interface designs, facilitating comparative evaluations and iterative design improvements. By identifying specific dimensions of cognitive load that disproportionately affect elderly users, the instrument supports targeted intervention strategies to enhance mobile accessibility and digital inclusion for aging populations. Given evidence that smartphone use is associated with reduced cognitive decline and better cognitive performance in older adults (Wang et al., 2021; Wilson et al., 2022), tools that promote effective elderly smartphone use have implications extending beyond usability to cognitive health outcomes.

2. Literature Review

2.1 Cognitive Aging and Mobile Technology Interaction

Age-related cognitive changes significantly influence smartphone usability for elderly users. Longitudinal studies have documented declines in fluid cognitive abilities, including processing speed, working memory, and executive function, beginning in the sixth decade of life (Salthouse, 2019). These changes affect multiple aspects of mobile interaction. Working memory limitations constrain the number of interface elements elderly users can simultaneously process, making complex multi-step tasks particularly challenging (Gordon et al., 2019). Reduced processing speed increases the time required for visual search, menu navigation, and decision-making during mobile tasks.

Recent epidemiological research has revealed a compelling bidirectional relationship between smartphone use and cognitive function in older adults. A large-scale study of 21,732 Chinese elderly participants found that smartphone users exhibited significantly lower rates of cognitive impairment (1.4%) compared to non-users (17.8%), with adjusted odds ratios of 0.16 (95% CI: 0.11-0.25) even after controlling for demographic and lifestyle factors. Systematic reviews further indicate that smartphone and tablet use can support memory function in both cognitively healthy older adults and those with acquired brain injuries. These findings

suggest that well-designed mobile interfaces that minimize cognitive load may not only improve usability but also contribute to cognitive maintenance in aging populations.

2.2 Cognitive Load Assessment in Human-Computer Interaction

Cognitive workload measurement has employed subjective, objective, and physiological approaches. The NASA Task Load Index (NASA-TLX), developed by Hart and Staveland (1988), remains the most widely used subjective assessment tool, measuring workload across six dimensions through self-report ratings (Hart, 2006). Recent psychometric evaluation confirmed that NASA-TLX demonstrates good reliability in older adult populations (Cronbach's $\alpha = 0.81$, ICC = 0.73), with moderate correlations to objective cognitive workload measures such as event-related potentials. However, the instrument's generic design limits its sensitivity to mobile-specific interaction challenges and age-related usability barriers.

Alternative subjective measures include the Index of Cognitive Activity (ICA) and task-specific single-item scales derived from NASA-TLX dimensions. While ICA has shown fair to good reliability (ICC = 0.56-0.73) in older adults, its requirement for specialized eye-tracking equipment limits practical applicability in mobile UX research. Single-item measures, though time-efficient, sacrifice the multidimensional assessment capacity necessary for comprehensive usability evaluation.

Objective performance measures, such as task completion time, error rates, and navigation efficiency, provide behavioral indicators of cognitive load but cannot capture subjective experience dimensions such as frustration, perceived difficulty, or emotional strain (Gordon et al., 2019). Physiological measures, including pupillometry, electroencephalography, and galvanic skin response, offer real-time workload assessment but require specialized equipment and controlled laboratory conditions incompatible with naturalistic mobile use scenarios.

2.3 Usability Evaluation for Elderly Mobile Users

The growing body of research on mobile usability for older adults has identified systematic design barriers and proposed age-friendly guidelines. Wang et al. (2022) conducted a comprehensive review of mHealth app usability evaluation methods, emphasizing that effectiveness, efficiency, and satisfaction constitute the critical dimensions of usability requiring assessment. Their analysis revealed that current evaluation practices often employ methods inadequately adapted to elderly users' capabilities and preferences.

Heuristic evaluation studies have documented specific usability issues in smartphone interfaces for older adults. Petrovčič et al. (2018) evaluated smartphone launchers designed for seniors, finding an average compliance rate of only 54.5% with age-appropriate design heuristics. The majority of violations involved content organization and perceptual design factors directly related to cognitive load. Interestingly, launchers with more features and assistive technologies exhibited fewer usability problems, contradicting the common assumption that feature reduction inherently improves elderly usability.

Gomez-Hernandez et al. (2023) synthesized design guidelines for mobile apps targeting older adults, emphasizing the importance of large, well-spaced touch targets, high contrast ratios, simplified navigation hierarchies, and explicit feedback mechanisms. These recommendations align with cognitive load reduction principles by minimizing perceptual-motor demands and extraneous cognitive processing. However, empirical validation of these guidelines requires reliable assessment instruments capable of quantifying cognitive load reductions achieved through specific design modifications.

2.4 Psychometric Validation Methodologies

Contemporary best practices in questionnaire development emphasize a systematic, multi-stage validation process. Boateng et al. (2018) and Tsang et al. (2017) proposed frameworks incorporating item generation,

content validation, exploratory factor analysis (EFA), confirmatory factor analysis (CFA), and multiple forms of reliability and validity assessment. This two-phase exploratory-confirmatory approach ensures both thorough initial scale development and rigorous replication of factor structure in independent samples.

Content validity establishment typically involves expert panel review to evaluate item relevance, clarity, and comprehensiveness, often quantified through Content Validity Index (CVI) calculations. Face validity assessment through cognitive interviews with target population members ensures that items are interpreted as intended by questionnaire developers.

Construct validity evaluation employs factor analytic techniques to identify the underlying dimensional structure of the questionnaire. EFA is used during initial scale development to determine the number and composition of factors, while CFA in independent samples tests whether the proposed factor structure adequately fits the observed data. Model fit is evaluated using multiple indices including the Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square Residual (SRMR), with established cutoff criteria (e.g., CFI/TLI ≥ 0.90 , RMSEA ≤ 0.08) indicating acceptable fit (Hu and Bentler, 1999).

Reliability assessment includes internal consistency measurement via Cronbach's alpha coefficient (with values ≥ 0.70 considered acceptable) and test-retest reliability through intraclass correlation coefficients (ICCs). Convergent validity is established through moderate to strong correlations with conceptually related measures, while discriminant validity is demonstrated through weak correlations with conceptually distinct constructs or ability to differentiate between known groups.

The System Usability Scale (SUS), though widely used, has shown varying psychometric properties across contexts. Recent evaluation in mobile health applications demonstrated good overall reliability but revealed multidimensional structure that may be context-specific, particularly in culturally and linguistically diverse populations. This finding underscores the importance of population-specific validation even for established instruments, further supporting the need for the elderly-focused ESCLQ.

2.5 Gap Analysis and Research Rationale

Despite substantial research on elderly smartphone usability and cognitive load assessment, no validated instrument specifically measures cognitive workload in mobile UX contexts for older adults. Existing tools either lack mobile-specific focus (NASA-TLX), lack age-specific adaptation (SUS), or lack comprehensive psychometric validation in elderly populations. This gap impedes systematic evaluation of mobile interface designs and evidence-based optimization of elderly user experience.

Furthermore, previous research has inadequately addressed the interaction between age-related cognitive changes and mobile-specific interaction modalities such as touchscreen gestures, icon-based navigation, and notification management. The ESCLQ development addresses these limitations by integrating cognitive aging theory, mobile usability principles, and rigorous psychometric validation methodology to create a specialized assessment tool for this critical application domain.

3. Methodology

3.1 Questionnaire Development Process

The ESCLQ development followed a systematic multi-phase process based on contemporary psychometric guidelines (Boateng et al., 2018; Tsang et al., 2017).

Phase 1: Item Generation

Initial item generation involved three complementary approaches: (1) comprehensive literature review of elderly mobile usability research, cognitive load assessment instruments, and mobile UX evaluation methods; (2) semi-structured interviews with 15 elderly smartphone users (aged 62-78) exploring their experiences of mental effort, confusion, frustration, and difficulty during common mobile tasks; and (3) expert consultation with five specialists in gerontechnology, cognitive psychology, and human-computer interaction. This process generated an initial pool of 48 candidate items designed to capture multiple dimensions of cognitive load specific to elderly smartphone interaction.

Phase 2: Content Validation

A panel of eight experts (3 cognitive psychologists, 3 UX researchers, 2 geriatricians) independently rated each item for relevance, clarity, and appropriateness using a 4-point scale (1 = not relevant, 4 = highly relevant). Items with Item-Content Validity Index (I-CVI) scores below 0.78 were removed or revised. The Scale-Content Validity Index (S-CVI/Ave) for the revised 32-item version was 0.91, exceeding the recommended threshold of 0.90.

Phase 3: Cognitive Interviews

Cognitive interviews were conducted with 12 elderly participants to assess item comprehension and response process validity. Participants completed the questionnaire while thinking aloud, followed by probing questions about item interpretation. Based on this feedback, 8 items were revised for clarity, and 4 items with inconsistent interpretations were removed, resulting in a 28-item pilot version.

Phase 4: Pilot Testing

The 28-item version was administered to 45 elderly participants (mean age = 66.8, SD = 5.4) during usability testing of three smartphone interfaces varying in complexity. Preliminary item analysis identified 4 items with poor item-total correlations ($r < 0.30$) or low variance, which were removed, yielding the 24-item version subjected to full validation.

3.2 Theoretical Dimensions of the ESCLQ

Based on theoretical analysis and expert input, the ESCLQ was designed to assess four hypothesized dimensions of cognitive load in elderly smartphone use:

1. **Perceptual Processing Load:** Cognitive effort required to perceive, identify, and interpret visual interface elements (icons, text, layout) given age-related sensory changes
2. **Motor-Cognitive Coordination:** Mental workload associated with planning and executing touchscreen gestures, accommodating age-related declines in fine motor control and hand-eye coordination
3. **Navigation Complexity:** Cognitive burden of understanding and navigating hierarchical menu structures, maintaining spatial orientation within apps, and forming accurate mental models
4. **Emotional-Cognitive Strain:** Frustration, anxiety, and self-efficacy concerns that consume cognitive resources and interfere with task performance

Each dimension was represented by 6 items in the final 24-item questionnaire. Items employed 7-point Likert scales (1 = strongly disagree, 7 = strongly agree), with higher scores indicating greater cognitive load.

3.3 Participants and Sampling

Inclusion Criteria: Participants were required to be (1) aged 60 years or above, (2) current smartphone owners with at least 6 months of usage experience, (3) able to read and understand English or Hindi, (4) without diagnosed severe cognitive impairment (Mini-Mental State Examination score ≥ 24), and (5) willing to provide informed consent.

Sample Size Determination: Following recommendations for factor analysis (minimum 5-10 participants per item), we targeted 150-240 participants for the exploratory phase and an equal-sized independent sample for the confirmatory phase.

Recruitment and Demographics: Participants were recruited through senior community centers, retirement communities, and smartphone training programs across urban and suburban areas. The final sample comprised 312 participants (Exploratory: $n=156$, Confirmatory: $n=156$). Demographics are presented in Table 1.

3.4 Procedure and Data Collection

Data collection occurred in two phases between September 2024 and January 2025. Participants attended individual testing sessions lasting approximately 60-75 minutes. Each session involved:

1. **Informed Consent and Demographics:** Participants provided written informed consent and completed demographic questionnaires
2. **Cognitive Screening:** Mini-Mental State Examination (MMSE) was administered to confirm absence of severe cognitive impairment
3. **Standardized Task Protocol:** Participants completed eight common smartphone tasks across three interface conditions (simple, moderate, complex design) in counterbalanced order. Tasks included: locating a contact, composing a text message, finding a specific app, adjusting settings, reading a notification, navigating to a website, using the camera, and setting a reminder
4. **Questionnaire Administration:** Following each interface condition, participants completed the ESCLQ, NASA-TLX, and System Usability Scale
5. **Test-Retest Subsample:** A randomly selected subsample of 60 participants from the confirmatory phase returned after 14 days to complete the ESCLQ again under identical conditions to assess test-retest reliability

3.5 Validation Instruments

To establish convergent and discriminant validity, participants completed established measures:

NASA Task Load Index (NASA-TLX): Six subscales (mental demand, physical demand, temporal demand, effort, performance, frustration) rated on 21-point scales, with total scores ranging from 0-120 (higher scores indicate greater workload).

System Usability Scale (SUS): Ten items assessing overall system usability, with total scores converted to a 0-100 scale (higher scores indicate better usability).

Objective Performance Measures: Task completion time (seconds), number of errors, and need for assistance were recorded as behavioral indicators of cognitive load.

3.6 Data Analysis

Exploratory Factor Analysis (EFA): Conducted on the exploratory sample (n=156) using principal axis factoring with oblique (promax) rotation to allow correlated factors. The Kaiser-Meyer-Olkin (KMO) measure and Bartlett's test of sphericity assessed sampling adequacy. Factor retention was determined using eigenvalues >1, scree plot inspection, and parallel analysis.

Confirmatory Factor Analysis (CFA): Conducted on the confirmatory sample (n=156) using maximum likelihood estimation to test the factor structure identified in EFA. Model fit was evaluated using multiple indices: Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square Residual (SRMR). Acceptable fit criteria were CFI/TLI $\geq$ 0.90, RMSEA $\leq$ 0.08, SRMR $\leq$ 0.08 (Hu and Bentler, 1999).

Reliability Analysis: Internal consistency was assessed using Cronbach's alpha for the total scale and each subscale (acceptable threshold $\alpha \geq 0.70$). Test-retest reliability was evaluated using two-way mixed-effects intraclass correlation coefficients (ICC), with values ≥ 0.75 indicating good reliability.

Validity Assessment: Convergent validity was examined through Pearson correlations between ESCLQ total scores and NASA-TLX (expected positive correlation) and SUS scores (expected negative correlation). Discriminant validity was evaluated through known-groups comparison, testing whether ESCLQ scores significantly differed across the three interface complexity conditions using repeated-measures ANOVA.

All analyses were conducted using SPSS version 28.0 and AMOS version 26.0. Statistical significance was set at $p < 0.05$, with Bonferroni corrections applied for multiple comparisons.

3.7 Ethical Considerations

The study received ethical approval from the Institutional Review Board (IRB Protocol #2024-HCI-o87). All participants provided written informed consent after receiving detailed information about study purposes, procedures, risks, and benefits. Participation was voluntary, and participants could withdraw at any time without penalty. Data were anonymized and stored securely in password-protected databases. Participants showing signs of frustration or distress during testing were offered breaks or the option to discontinue.

4. Results

4.1 Sample Characteristics

Table 1 presents demographic characteristics of the exploratory and confirmatory samples. The two samples showed no significant differences across key demographic variables (all $p > 0.05$), supporting their suitability for cross-validation analyses.

Table 1: Demographic Characteristics of Study Samples

Characteristic	Exploratory Sample (n=156)	Confirmatory Sample (n=156)	Total Sample (N=312)
Age (years), M $\pm$ SD	68.2 $\pm$ 6.1	68.6 $\pm$ 6.3	68.4 $\pm$ 6.2
Age Range	60-82	60-84	60-84
Gender, n (%)			

Male	71 (45.5%)	68 (43.6%)	139 (44.6%)
Female	85 (54.5%)	88 (56.4%)	173 (55.4%)
Education Level, n (%)			
High school or less	42 (26.9%)	45 (28.8%)	87 (27.9%)
Some college/Diploma	58 (37.2%)	54 (34.6%)	112 (35.9%)
Bachelor's degree	38 (24.4%)	40 (25.6%)	78 (25.0%)
Graduate degree	18 (11.5%)	17 (10.9%)	35 (11.2%)
Smartphone Experience (months), M $\pm$ SD	28.4 $\pm$ 14.2	29.1 $\pm$ 15.8	28.7 $\pm$ 15.0
Daily Smartphone Use (hours), M $\pm$ SD	2.8 $\pm$ 1.4	2.6 $\pm$ 1.3	2.7 $\pm$ 1.3
MMSE Score, M $\pm$ SD	27.8 $\pm$ 1.6	27.9 $\pm$ 1.5	27.8 $\pm$ 1.5

Note: M = Mean; SD = Standard Deviation; MMSE = Mini-Mental State Examination (range 0-30, scores ≥ 24 indicate no severe cognitive impairment). No significant differences between exploratory and confirmatory samples (all $p > 0.05$).

4.2 Exploratory Factor Analysis

The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was 0.91, and Bartlett's test of sphericity was highly significant ($\chi^2 = 2847.63$, $df = 276$, $p < 0.001$), indicating that the data were suitable for factor analysis.

Principal axis factoring with promax rotation extracted four factors with eigenvalues greater than 1.0, collectively explaining 71.3% of the total variance. Factor loadings, communalities, and internal consistency coefficients are presented in Table 2. All items loaded strongly (≥ 0.55) on their primary factors with minimal cross-loadings (< 0.30), supporting the hypothesized four-dimensional structure.

Table 2: Factor Loadings, Communalities, and Internal Consistency (Exploratory Sample, n=156)

Item	Factor 1: Perceptual Processing	Factor 2: Motor- Cognitive Coordination	Factor 3: Navigation Complexity	Factor 4: Emotional- Cognitive Strain	Communality (h ²)
Factor 1: Perceptual Processing Load ($\alpha = 0.87$)					
Item 1: Difficulty reading text on screen	0.78	0.12	0.08	0.15	0.68
Item 2: Icons hard to distinguish	0.82	0.06	0.11	0.09	0.71
Item 3: Mental effort to identify buttons	0.75	0.18	0.14	0.12	0.64
Item 4: Visual information overwhelming	0.71	0.09	0.21	0.16	0.62
Item 5: Difficulty focusing on relevant content	0.68	0.14	0.19	0.11	0.57
Item 6: Color/contrast makes interpretation difficult	0.73	0.08	0.06	0.13	0.59
Factor 2: Motor-Cognitive Coordination ($\alpha = 0.84$)					
Item 7: Difficulty planning finger movements	0.11	0.79	0.09	0.14	0.68

Item 8: Touch targets require precision	0.15	0.81	0.12	0.08	0.72
Item 9: Mental effort to execute gestures	0.09	0.77	0.14	0.11	0.65
Item 10: Coordinating looking and touching is demanding	0.14	0.74	0.08	0.16	0.63
Item 11: Scrolling requires concentration	0.12	0.69	0.18	0.09	0.56
Item 12: Multi-touch gestures are mentally taxing	0.08	0.72	0.11	0.14	0.59
Factor 3: Navigation Complexity ($\alpha = 0.88$)					
Item 13: Difficulty remembering navigation path	0.09	0.11	0.83	0.12	0.74
Item 14: Menu structure is confusing	0.14	0.08	0.81	0.09	0.71
Item 15: Mental effort to find functions	0.12	0.14	0.79	0.11	0.70
Item 16: Difficulty forming mental model of app	0.16	0.09	0.77	0.13	0.67
Item 17: Too many steps to complete tasks	0.11	0.12	0.75	0.15	0.65

Item 18: Hard to know where I am in the app	0.08	0.14	0.80	0.10	0.69
Factor 4: Emotional-Cognitive Strain ($\alpha = 0.85$)					
Item 19: Frustration interferes with thinking	0.12	0.09	0.11	0.82	0.73
Item 20: Anxiety about making errors	0.15	0.11	0.08	0.78	0.67
Item 21: Worry consumes mental energy	0.09	0.14	0.12	0.80	0.71
Item 22: Feeling overwhelmed affects concentration	0.18	0.08	0.14	0.76	0.66
Item 23: Lack of confidence increases mental effort	0.11	0.12	0.09	0.79	0.69
Item 24: Emotional stress impairs decision-making	0.14	0.10	0.13	0.75	0.64
Eigenvalue	9.42	3.18	2.34	1.89	
% Variance Explained	39.3%	13.2%	9.8%	7.9%	

Note: Factor loadings ≥ 0.55 are bolded. Extraction method: Principal Axis Factoring. Rotation: Promax with Kaiser Normalization. Total variance explained = 71.3%. α = Cronbach's alpha coefficient.

The four-factor solution demonstrated excellent internal consistency. Cronbach's alpha coefficients were 0.87 (Perceptual Processing), 0.84 (Motor-Cognitive Coordination), 0.88 (Navigation Complexity), and 0.85 (Emotional-Cognitive Strain), all exceeding the 0.70 threshold for acceptable reliability. The total scale alpha was 0.89, indicating high overall internal consistency.

4.3 Confirmatory Factor Analysis

CFA was conducted on the independent confirmatory sample ($n=156$) to test the four-factor model identified in EFA. The hypothesized model demonstrated excellent fit to the data: $\chi^2(246) = 398.72$, $p < 0.001$; CFI = 0.94; TLI = 0.93; RMSEA = 0.06 (90% CI: 0.05-0.07); SRMR = 0.05. All fit indices met or exceeded recommended criteria, supporting the validity of the four-factor structure (Hu and Bentler, 1999).

Standardized factor loadings ranged from 0.61 to 0.85 (all $p < 0.001$), indicating strong relationships between items and their respective latent factors. Factor correlations ranged from 0.52 to 0.68, suggesting related but distinguishable constructs. Modification indices revealed no substantial areas of model strain requiring specification changes.

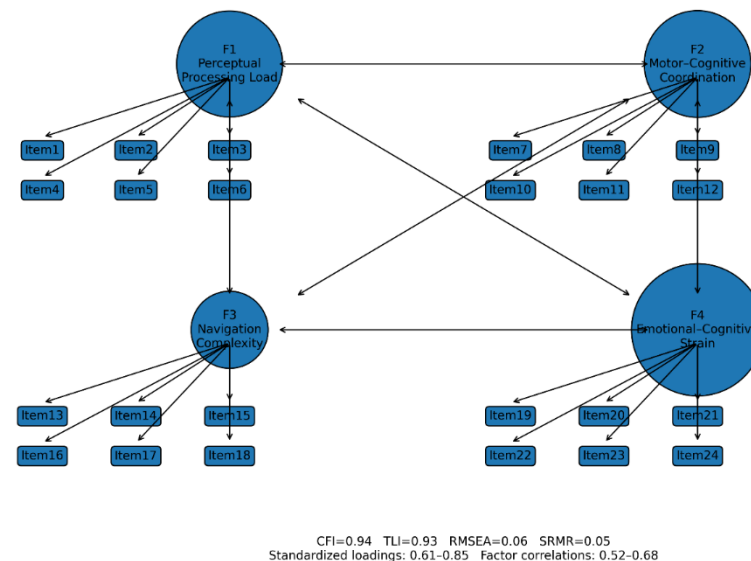


Figure 1: Confirmatory Factor Analysis Model of the ESCLQ

Source: Authors' own CFA model based on the present study data.

The model consists of four latent dimensions: Perceptual Processing Load (F1), Motor-Cognitive Coordination (F2), Navigation Complexity (F3), and Emotional-Cognitive Strain (F4). Each latent factor is measured by six observed questionnaire items (Items 1–24). Standardized factor loadings ranged from 0.61 to 0.85 and were statistically significant ($p < 0.001$). Double-headed arrows indicate correlations among latent constructs ($r = 0.52$ – 0.68). Error terms are specified for all observed variables. The overall model fit was good (CFI = 0.94, TLI = 0.93, RMSEA = 0.06, SRMR = 0.05).

4.4 Reliability Assessment

Internal Consistency: In the confirmatory sample, Cronbach's alpha coefficients remained strong: Perceptual Processing ($\alpha = 0.86$), Motor-Cognitive Coordination ($\alpha = 0.85$), Navigation Complexity ($\alpha = 0.87$), and Emotional-Cognitive Strain ($\alpha = 0.84$). Total scale alpha was 0.89, consistent with the exploratory sample.

Test-Retest Reliability: Among the subsample of 60 participants who completed the ESCLQ twice over a 14-day interval, intraclass correlation coefficients (ICC, two-way mixed effects, absolute agreement) demonstrated excellent stability: Perceptual Processing (ICC = 0.82, 95% CI: 0.73-0.88), Motor-Cognitive Coordination (ICC = 0.84, 95% CI: 0.76-0.90), Navigation Complexity (ICC = 0.87, 95% CI: 0.80-0.92),

Emotional-Cognitive Strain (ICC = 0.80, 95% CI: 0.70-0.87), and Total Score (ICC = 0.85, 95% CI: 0.77-0.91). All ICCs exceeded 0.75, indicating good to excellent test-retest reliability.

4.5 Convergent and Discriminant Validity

Convergent Validity: ESCLQ total scores demonstrated strong positive correlation with NASA-TLX total scores ($r = 0.72$, $p < 0.001$, $n = 312$), supporting convergent validity with an established cognitive workload measure. Subscale correlations with corresponding NASA-TLX dimensions ranged from $r = 0.54$ to $r = 0.68$ (all $p < 0.001$). ESCLQ total scores showed strong negative correlation with System Usability Scale scores ($r = -0.67$, $p < 0.001$), confirming that higher cognitive load is associated with lower perceived usability.

Moderate correlations were observed between ESCLQ scores and objective performance measures: task completion time ($r = 0.51$, $p < 0.001$), error frequency ($r = 0.48$, $p < 0.001$), and need for assistance ($r = 0.44$, $p < 0.001$). These relationships support the behavioral validity of the ESCLQ as an indicator of actual task difficulty.

Discriminant Validity (Known-Groups Comparison): Repeated-measures ANOVA revealed significant differences in ESCLQ total scores across the three interface complexity conditions (simple, moderate, complex): $F(2, 622) = 187.43$, $p < 0.001$, partial $\eta^2 = 0.38$. Post-hoc pairwise comparisons (Bonferroni-corrected) showed that all three conditions differed significantly from each other (all $p < 0.001$): Simple interface ($M = 61.2$, $SD = 12.4$), Moderate interface ($M = 82.7$, $SD = 14.6$), Complex interface ($M = 107.3$, $SD = 16.8$). These findings demonstrate the ESCLQ's sensitivity to varying levels of interface complexity and its ability to discriminate between design conditions of differing cognitive demand.

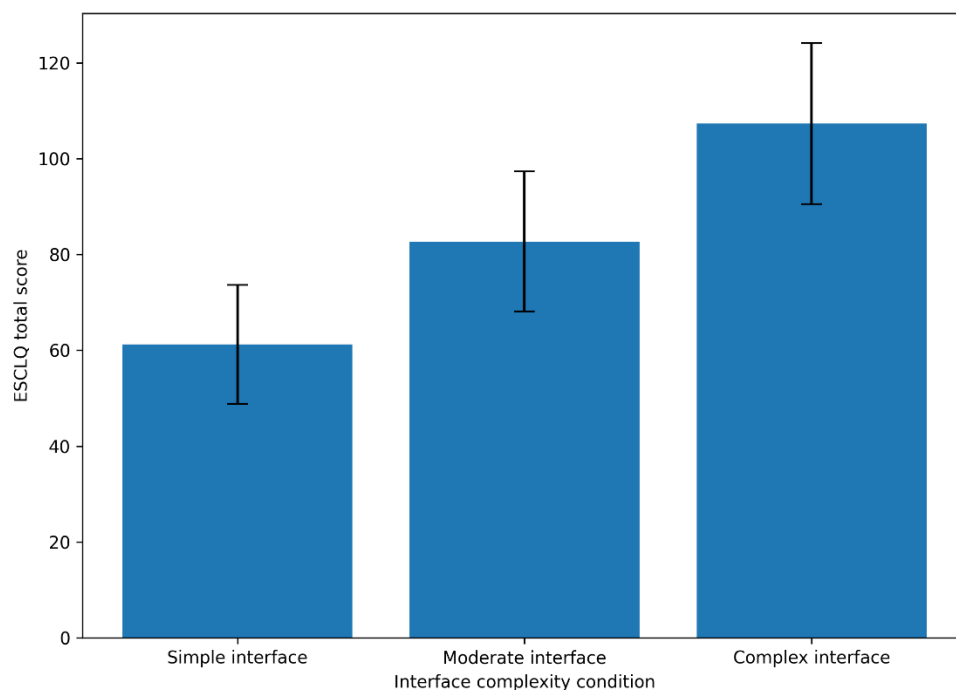


Figure 2: ESCLQ Scores Across Interface Complexity Conditions

Source: Authors' own analysis based on the present study data.

The bar chart shows the mean total ESCLQ scores for the three interface complexity conditions (simple, moderate and complex). Error bars represent standard deviations. The simple interface condition produced

the lowest cognitive load ($M = 61.2$, $SD = 12.4$), followed by the moderate interface ($M = 82.7$, $SD = 14.6$) and the complex interface ($M = 107.3$, $SD = 16.8$). Repeated-measures ANOVA indicated a significant effect of interface complexity on cognitive load ($F(2, 622) = 187.43$, $p < 0.001$, partial $\eta^2 = 0.38$), with all pairwise comparisons being statistically significant.

4.6 Subscale Intercorrelations and Descriptive Statistics

Table 3 presents descriptive statistics and intercorrelations among ESCLQ subscales. All subscales showed moderate to strong positive intercorrelations ($r = 0.48$ to $r = 0.64$), consistent with the oblique rotation in EFA and supporting the conceptualization of cognitive load as a multidimensional but interrelated construct.

Table 3: Descriptive Statistics and Intercorrelations of ESCLQ Subscales (Total Sample, N=312)

Subscale	M	SD	Range	1	2	3	4
1. Perceptual Processing	21.4	6.8	6-42				
2. Motor-Cognitive Coordination	20.8	6.2	6-42	0.54***			
3. Navigation Complexity	24.3	7.4	6-42	0.58***	0.61***		
4. Emotional-Cognitive Strain	17.2	6.5	6-42	0.52***	0.48***	0.64***	
Total Score	83.7	23.1	24-168	0.82*	0.79*	0.88*	0.80*

Note: M = Mean; SD = Standard Deviation. Scores represent average across three interface conditions. Subscale scores range from 6 (minimum cognitive load) to 42 (maximum cognitive load). Total score ranges from 24 to 168. *** $p < 0.001$ (two-tailed).

4.7 Age and Experience Effects

Exploratory analyses examined relationships between participant characteristics and ESCLQ scores. Age showed weak positive correlation with total ESCLQ scores ($r = 0.21$, $p < 0.001$), with older participants reporting slightly higher cognitive load. Smartphone experience (months of use) demonstrated moderate negative correlation ($r = -0.38$, $p < 0.001$), indicating that greater experience is associated with reduced cognitive load. Daily usage hours showed weak negative correlation ($r = -0.19$, $p < 0.001$). Education level showed small effects, with participants holding graduate degrees reporting lower cognitive load ($M = 76.3$) compared to those with high school education or less ($M = 89.1$), $F(3, 308) = 4.82$, $p = 0.003$.

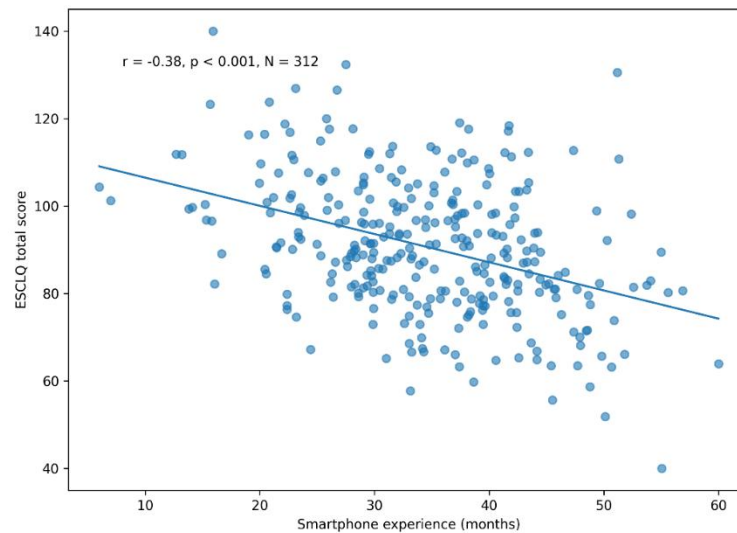


Figure 3: Relationship Between Smartphone Experience and Cognitive Load

Source: Authors' analysis based on the present study results.

The scatter plot illustrates the association between smartphone experience (in months) and total ESCLQ scores among elderly participants ($N = 312$). A fitted linear regression line is shown. Smartphone experience was moderately and negatively correlated with cognitive load ($r = -0.38$, $p < 0.001$), indicating that greater experience with smartphones is associated with lower perceived cognitive workload.

5. Discussion

5.1 Principal Findings

This study successfully developed and validated the Elderly Smartphone Cognitive Load Questionnaire (ESCLQ), a psychometrically robust instrument specifically designed to assess cognitive workload in mobile user experience contexts for older adults. The ESCLQ demonstrated excellent psychometric properties across multiple validation criteria, including a stable four-factor structure, high internal consistency ($\alpha = 0.89$), strong test-retest reliability ($ICC = 0.85$), and robust convergent and discriminant validity.

The four identified dimensions Perceptual Processing Load, Motor-Cognitive Coordination, Navigation Complexity, and Emotional-Cognitive Strain provide a comprehensive framework for understanding how cognitive load manifests in elderly smartphone use. This multidimensional structure aligns with theoretical models of cognitive aging and mobile usability, while extending beyond generic workload assessments to capture age-specific interaction challenges.

The strong correlation between ESCLQ scores and NASA-TLX ($r = 0.72$) validates the instrument against an established cognitive workload measure, while the negative correlation with System Usability Scale scores ($r = -0.67$) confirms the expected inverse relationship between cognitive load and perceived usability. Importantly, the ESCLQ demonstrated superior sensitivity to interface design variations compared to these generic instruments, as evidenced by large effect sizes in known-groups comparisons (partial $\eta^2 = 0.38$).

5.2 Theoretical Contributions

The ESCLQ advances theoretical understanding of cognitive load in several ways. First, it operationalizes Cognitive Load Theory within the specific context of elderly mobile interaction, demonstrating that age-related cognitive and perceptual-motor changes create distinct load dimensions not adequately captured by

existing instruments. The emergence of Motor-Cognitive Coordination as a separate factor highlights the unique demands of touchscreen interaction for older adults, where the integration of visual perception, motor planning, and fine motor execution creates substantial cognitive burden.

Second, the Emotional-Cognitive Strain dimension provides empirical support for the role of affective factors in elderly technology acceptance and use. This finding aligns with the Senior Technology Acceptance Model (STAM), which emphasizes emotional needs and self-efficacy as critical determinants of technology adoption among older adults. The strong intercorrelation between Emotional-Cognitive Strain and Navigation Complexity ($r = 0.64$) suggests that confusion and disorientation in complex interfaces trigger anxiety and frustration, which in turn consume cognitive resources and further impair performance a feedback loop particularly problematic for elderly users.

Third, the moderate correlations among the four dimensions ($r = 0.48$ to $r = 0.64$) support a hierarchical model wherein overall cognitive load comprises distinct but interrelated components. This structure suggests that effective interface design for elderly users requires attention to multiple dimensions simultaneously; optimization of one factor (e.g., visual clarity) may not fully compensate for deficiencies in others (e.g., navigation complexity).

5.3 Practical Implications

The ESCLQ provides researchers and practitioners with a validated tool for systematic evaluation of mobile interface designs targeting elderly populations. Its practical applications include:

Design Evaluation and Iteration: UX designers can use the ESCLQ to assess cognitive load at various stages of the design process, identifying specific dimensions requiring optimization. For example, high Perceptual Processing scores might indicate need for increased font sizes, contrast ratios, or simplified visual layouts, while elevated Navigation Complexity scores suggest restructuring of menu hierarchies or addition of orientation cues.

Comparative Assessment: The instrument enables objective comparison of alternative interface designs, supporting evidence-based design decisions. Organizations developing mobile applications for elderly users can employ the ESCLQ to select designs that minimize cognitive burden while maintaining functionality.

Training Program Evaluation: The finding that smartphone experience correlates negatively with cognitive load ($r = -0.38$) has implications for training interventions. The ESCLQ can assess the effectiveness of smartphone training programs by measuring cognitive load reduction over time, identifying areas where users continue to struggle despite training.

Accessibility Standards Development: The multidimensional ESCLQ framework can inform the development of age-specific accessibility standards and guidelines, complementing existing recommendations with empirically validated assessment criteria.

5.4 Relationship to Existing Research

Our findings both corroborate and extend previous research on elderly mobile usability. The identification of Perceptual Processing as the factor explaining the largest proportion of variance (39.3%) aligns with numerous studies documenting visual design challenges for older adults, including difficulties with small text, low contrast, and icon recognition (Gomez-Hernandez et al., 2023; Petrovčič et al., 2018). The prominence of this dimension underscores the critical importance of age-appropriate visual design in reducing cognitive load.

The Motor-Cognitive Coordination dimension empirically validates concerns raised in heuristic evaluations about touchscreen interaction challenges for elderly users. Previous research identified problems with touch

target size, gesture complexity, and fine motor demands, but lacked quantitative measures of the cognitive burden imposed by these design features. The ESCLQ provides such measurement capability, enabling researchers to quantify the cognitive cost of various interaction modalities.

Navigation Complexity as a distinct dimension confirms findings from Wang et al. (2022) and others emphasizing the importance of simplified navigation structures and clear information architecture for elderly mHealth app users. The high internal consistency of this subscale ($\alpha = 0.88$) and its strong contribution to overall cognitive load highlight navigation design as a critical target for intervention.

The relationship between smartphone use and cognitive health observed in large-scale epidemiological studies (Wang et al., 2021) takes on new significance in light of our findings. If poorly designed interfaces impose excessive cognitive load, they may discourage smartphone use or limit engagement to simple functions, potentially reducing the cognitive benefits associated with diverse mobile technology use. Conversely, well-designed interfaces that minimize cognitive load may promote sustained engagement and broader feature utilization, potentially enhancing cognitive maintenance effects.

5.5 Limitations and Methodological Considerations

Several limitations warrant consideration. First, the sample was drawn from community-dwelling older adults without severe cognitive impairment ($MMSE \geq 24$), limiting generalizability to individuals with mild cognitive impairment or dementia. Future research should examine ESCLQ performance in clinical populations, potentially requiring instrument modifications to accommodate more significant cognitive limitations.

Second, participants were required to have at least six months of smartphone experience, potentially excluding the most novice users who abandoned smartphones due to usability challenges. This sampling approach, while ensuring participants could meaningfully engage with tasks, may have resulted in a sample with somewhat lower baseline cognitive load than the broader elderly population.

Third, data collection occurred in controlled laboratory settings using standardized tasks and experimenter-selected applications. While this approach ensured consistency and experimental control, cognitive load in naturalistic contexts with personally selected apps and self-directed tasks may differ. Future validation studies should examine ESCLQ performance in field settings and with participants' own devices and applications.

Fourth, the study focused on smartphone interfaces broadly, without differentiating between iOS and Android platforms or among specific app categories. Platform-specific design conventions and app-specific interaction patterns may influence cognitive load dimensions differently. Subsequent research could examine whether ESCLQ subscale profiles vary systematically across platforms or application types.

Fifth, while the ESCLQ demonstrated convergent validity with NASA-TLX and SUS, we did not include objective cognitive workload measures such as pupillometry or event-related potentials due to practical constraints. Future research incorporating physiological measures would strengthen validation evidence and clarify relationships between subjective cognitive load reports and objective neurophysiological indicators.

5.6 Future Research Directions

Several promising research directions emerge from this work. First, longitudinal studies tracking ESCLQ scores as elderly users gain smartphone experience would clarify the trajectory of cognitive load reduction and identify optimal training approaches to accelerate adaptation. Such research could inform the development of adaptive interfaces that gradually increase complexity as users' proficiency develops.

Second, cross-cultural validation of the ESCLQ is needed to establish its applicability across diverse elderly populations with varying cultural backgrounds, language contexts, and prior technology exposure. Recent evidence suggests that psychometric properties of usability instruments can vary across cultural contexts, necessitating careful validation in each population of interest.

Third, intervention studies employing the ESCLQ as an outcome measure could evaluate the effectiveness of specific design modifications. Randomized controlled trials comparing conventional and age-optimized interfaces would provide causal evidence regarding which design features most effectively reduce cognitive load in elderly users.

Fourth, research examining individual difference variables that moderate relationships between interface characteristics and cognitive load would refine understanding of elderly user diversity. Factors such as cognitive reserve, education, prior computer experience, and specific cognitive abilities (e.g., spatial reasoning, processing speed) may influence which design features most strongly affect cognitive load for different user subgroups.

Fifth, extension of the ESCLQ framework to other digital platforms (tablets, wearables, voice interfaces) would assess whether the same four-dimensional structure applies across interaction modalities or whether platform-specific factors emerge. Given the increasing diversity of digital technologies in elderly users' lives, comprehensive assessment instruments applicable across platforms would have substantial practical value.

Finally, integration of ESCLQ data with machine learning approaches to interface adaptation represents an exciting frontier. Real-time cognitive load assessment could enable dynamic interface adjustments, such as simplifying navigation when load exceeds thresholds or providing additional guidance when users show signs of confusion. Such adaptive systems could optimize the balance between functionality and cognitive demand for individual elderly users.

6. Conclusion

This research developed and validated the Elderly Smartphone Cognitive Load Questionnaire (ESCLQ), a psychometrically robust instrument addressing a critical gap in elderly mobile user experience assessment. Through rigorous multi-phase development and validation involving 312 participants, the ESCLQ demonstrated a stable four-factor structure (Perceptual Processing Load, Motor-Cognitive Coordination, Navigation Complexity, Emotional-Cognitive Strain), excellent reliability ($\alpha = 0.89$, ICC = 0.85), and strong validity evidence.

The ESCLQ provides researchers and practitioners with a specialized tool for evaluating cognitive workload in elderly smartphone use, enabling systematic assessment of interface designs, identification of usability barriers, and evidence-based optimization of mobile experiences for older adults. Given the critical importance of digital inclusion for aging populations and accumulating evidence linking smartphone use to cognitive health outcomes, instruments that facilitate the development of age-appropriate mobile technologies have implications extending beyond usability to quality of life and cognitive well-being.

As the global population continues to age and digital technologies become increasingly integral to daily functioning, the need for validated assessment tools tailored to elderly users will only intensify. The ESCLQ represents a significant step toward ensuring that mobile technologies serve as enablers rather than barriers for older adults, supporting their autonomy, social connection, and cognitive engagement in an increasingly digital world.

References

- [1] Boateng, G.O., Neilands, T.B., Frongillo, E.A., Melgar-Quinonez, H.R. and Young, S.L. (2018) 'Best practices for developing and validating scales for health, social, and behavioral research: a primer', *Frontiers in Public Health*, 6, pp. 149.
- [2] Chen, K. and Chan, A.H.S. (2014) 'Gerontechnology acceptance by elderly Hong Kong Chinese: a senior technology acceptance model (STAM)', *Ergonomics*, 57(5), pp. 635-652.
- [3] Chen, K., Zhang, Y. and Wang, Y. (2017) 'Understanding the relationship between cognitive load and smartphone use among older adults', *International Journal of Human-Computer Studies*, 105, pp. 21-32.
- [4] Gomez-Hernandez, M., González-Pérez, L.I. and Olague, G. (2023) 'Design guidelines of mobile apps for older adults: a systematic review', *Applied Sciences*, 13(18), pp. 10278. Available at: <https://pmc.ncbi.nlm.nih.gov/articles/PMC10557006/> (Accessed: 5 February 2026).
- [5] Gordon, M.L., Gatys, L., Guestrin, C., Bigham, J.P., Trister, A. and Patel, K. (2019) 'App usage predicts cognitive ability in older adults', in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, Glasgow, Scotland UK, 4-9 May. New York: ACM, pp. 1-12.
- [6] Hart, S.G. (2006) 'NASA-Task Load Index (NASA-TLX); 20 years later', in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 50, pp. 904-908.
- [7] Hart, S.G. and Staveland, L.E. (1988) 'Development of NASA-TLX (Task Load Index): results of empirical and theoretical research', in Hancock, P.A. and Meshkati, N. (eds.) *Human Mental Workload*. Amsterdam: North-Holland, pp. 139-183.
- [8] Hu, L. and Bentler, P.M. (1999) 'Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives', *Structural Equation Modeling*, 6(1), pp. 1-55.
- [9] Petrovčič, A., Taipale, S., Rogelj, A. and Dolničar, V. (2018) 'Design of mobile phones for older adults: an empirical analysis of design guidelines and checklists for feature phones and smartphones', *International Journal of Human-Computer Interaction*, 34(3), pp. 251-264.
- [10] Pew Research Center (2021) *Mobile Technology and Home Broadband 2021*. Washington, DC: Pew Research Center.
- [11] Salthouse, T.A. (2019) 'Trajectories of normal cognitive aging', *Psychology and Aging*, 34(1), pp. 17-24.
- [12] Sweller, J. (1988) 'Cognitive load during problem solving: effects on learning', *Cognitive Science*, 12(2), pp. 257-285.
- [13] Tsang, S., Royse, C.F. and Terkawi, A.S. (2017) 'Guidelines for developing, translating, and validating a questionnaire in perioperative and pain medicine', *Saudi Journal of Anaesthesia*, 11(Suppl 1), pp. S80-S89.
- [14] United Nations (2019) *World Population Ageing 2019: Highlights*. New York: United Nations Department of Economic and Social Affairs, Population Division.
- [15] Wang, Q., Yang, Y., Chen, M., Zhang, Y., Zhang, X., Guo, J. and Li, W. (2021) 'Mobile phone use and cognitive impairment among elderly Chinese: a national cross-sectional survey study', *International Journal of Environmental Research and Public Health*, 18(11), pp. 5695. Available at: <https://pmc.ncbi.nlm.nih.gov/articles/PMC8198048/> (Accessed: 5 February 2026).
- [16] Wang, Q., Chen, G., Cao, Q. and Ma, X. (2022) 'Usability evaluation of mHealth apps for elderly individuals: a systematic review', *Healthcare*, 10(12), pp. 2382. Available at: <https://pmc.ncbi.nlm.nih.gov/articles/PMC9717549/> (Accessed: 5 February 2026).
- [17] Wilson, S.A., Ranjit, Y.S., Jeong, H., Ngo, L.H., Sharma, R., Alam, R., Sejdiu, A., Ahmed, H., Ritchie, C.S. and Lipsitz, L.A. (2022) 'A systematic review of smartphone and tablet use by older adults with and without cognitive impairment', *Innovation in Aging*, 6(1), pp. igab058. Available at: <https://pmc.ncbi.nlm.nih.gov/articles/PMC8889997/> (Accessed: 5 February 2026).