Original Article
http://mjiri.iums.ac.ir
Medical Journal of the Islamic Republic of Iran (MJIRI)

Med J Islam Repub Iran. 2026 (13 Apr);40.35. https://doi.org/10.47176/mjiri.40.35

Evaluating the Psychometric Properties of the Persian Version of the Saint Louis University Mental Status (SLUMS) Exam Among a Group of Community-Dwelling Older Adults in Tehran, Iran

Shohreh Naderimagham[1,2], Farshad Sharifi[1,3], Sahar Keyvanloo Shahrestanaki[4], Hossein Fakhrzadeh[2,1], Zahra Amrollah Majdabadi Kohne[5], Fatemeh Ghonoodi[6], Sedigheh Sarabi[7], Negin Malekraeisi[8], Masoud Ghassemi[9]*, Masoud Arzaghi[1]*

Received: 16 Apr 2025 Accepted: 15 Mar 2026 Published: 13 Apr 2026

Abstract

Background: Early detection of cognitive impairments is crucial for preventing progression to dementia. An effective diagnosis necessitates a tool with robust diagnostic characteristics. The Saint Louis University Mental Status (SLUMS) exam is a widely utilized instrument for diagnosing neurocognitive disorders. This study assessed the psychometric properties of its Persian version (P-SLUMS) among Iranian community.

Methods: This cross-sectional study was conducted with 450 community-dwelling older adults who were referred to one of the Tehran Municipality centers and participated in a cultural program from March to November 2021. Data collection tools included a demographic questionnaire, the Mini-Mental State Examination (MMSE) for concurrent validity, and the SLUMS exam. Dementia was assessed using the Persian version of the Global Deterioration Scale (GDS), while depression was evaluated with the Persian version of the Patient Health Questionnaire-9 (PHQ-9). The SLUMS exam was translated into Persian in accordance with the World Health Organization's protocol. Participants were recruited through convenience sampling based on the inclusion criteria. The P-SLUMS exam was evaluated for content validity, face validity, concurrent validity, and construct validity using Confirmatory Factor Analysis (CFA), as well as reliability, which was assessed over a two-week interval among 20 participants.

Results: CFA indicated acceptable model fit indices. The goodness-of-fit indices for the model, including CFI, PCFI, and PNFI, were 0.82, 0.66, and 0.44, respectively, indicating a moderate fit that falls below the recommended thresholds for excellent parsimony. However, these indices are considered acceptable for our sample, which encompasses diverse education levels and a moderate number of items. The tool demonstrated high internal consistency (Cronbach's alpha = 0.866) and excellent test-retest reliability (ICC = 0.979).

Conclusion: The P-SLUMS exam is a valid, reliable, and user-friendly tool for early detection of Mild Cognitive Impairment (MCI) and dementia.

Keywords: SLUMS, Mild Cognitive Impairment, MCI, dementia, psychometric properties

Cite this article as: Naderimagham S, Sharifi F, Keyvanloo Shahrestanaki S, Fakhrzadeh H, Amrollah Majdabadi Kohne Z, Ghonoodi F, Sarabi S, Malekraeisi N, Ghassemi M, Arzaghi M. Evaluating the Psychometric Properties of the Persian Version of the Saint Louis University Mental Status (SLUMS) Exam Among a Group of Community-Dwelling Older Adults in Tehran, Iran. *Med J Islam Repub Iran.* 2026 (13 Apr);40:35. https://doi.org/10.47176/mjiri.40.35

Corresponding author: Dr Masoud Ghassemi, Mas_qasemi@iauctb.ac.ir
Dr Masoud Arzaghi, dr.arzaghi@gmail.com

1. Elderly Health Research Center, Endocrinology and Metabolism Population Sciences Institute, Tehran University of Medical Sciences, Tehran, Iran
2. Endocrinology and Metabolism Research Center, Endocrinology and Metabolism Clinical Sciences Institute, Tehran University of Medical Sciences, Tehran, Iran
3. Department of Gerontology, School of Rehabilitation, Tehran University of Medical Sciences, Tehran, Iran
4. Department of Community Health and Geriatric Nursing, Nursing and Midwifery Care Research Center, School of Nursing and Midwifery, Tehran University of Medical Sciences, Tehran, Iran
5. Department of Community Health and Geriatric Nursing, School of Nursing and Midwifery, Tehran University of Medical Sciences, Tehran, Iran
6. School of Medical, Shahid Beheshti University of Medical Sciences, Tehran, Iran
7. University of UKM, KL, Malaysia
8. Faculty of Medicine, Tehran Medical Science, Islamic Azad University, Tehran, Iran
9. Faculty of Educational Science and Psychology, Islamic Azad University, Central Tehran Branch, Tehran, Iran

↑What is "already known" in this topic:
Cognitive impairment (CI) and dementia are often underdiagnosed in early stages due to limitations of screening tools. Existing tools, such as the MMSE and MoCA, have accessibility and cost issues. The SLUMS exam, with higher sensitivity, greater educational adaptability, and free access, has shown promise across languages and populations.

→What this article adds:
This study provides the first psychometric evaluation of the Persian version of the SLUMS exam. It confirms its validity and reliability among older Persian-speaking adults, offering a culturally adapted, cost-free, and accurate tool for early cognitive impairment screening, particularly in resource-limited settings, such as Iran.

Introduction

Cognitive impairment represents a significant public health concern that is often undetectable in its early stages (1) and leads to a higher rate of hospitalization among the elderly compared to other neurocognitive disorders (2).

One significant factor contributing to the delay in the early detection of MCI and dementia is the limitations of existing diagnostic tools. An effective cognitive tool for detecting MCI and dementia should possess robust psychometric properties, be user-friendly, concise, and adequately assess cognitive functions (3). Although several tools have been developed for screening cognitive impairments, particularly in their early stages, only a few can comprehensively evaluate essential cognitive domains such as attention, executive function, orientation, memory, and language (4). Among these tools, the Montreal Cognitive Assessment (MoCA) and the MMSE are frequently employed, demonstrating relatively high sensitivity and specificity for detecting dementia (5–7).

Both tools exhibit two notable limitations. First, they are not suitable for older adults with limited literacy. Second, their purchase is necessary, which many health institutions in developing countries cannot afford (8).

The Saint Louis University Mental Status (SLUMS) exam is a screening tool for cognitive impairments developed in 2006 by the Saint Louis University School of Medicine Division. It consists of an 11-question, 30-point assessment that evaluates attention, executive function, memory, and orientation. The developers of this tool reported score ranges of 25 to 30 for normal cognitive function, 20 to 24 for MCI, and less than 20 for dementia (9, 10). The SLUMS exam takes approximately 15 to 20 minutes to complete and is more accurate than the MMSE for the early detection of cognitive impairments, which can be influenced by educational background. Several studies have demonstrated that the SLUMS exam detects mild cognitive impairment (MCI) with greater sensitivity than the MMSE, particularly during the early stages of cognitive decline. For instance, Tariq et al. (2006) reported that the SLUMS exam identified 63% of MCI cases, compared with 17% detected by the MMSE in a sample of older adults (11). Consistent with these findings, our study found a strong correlation between the P-SLUMS exam and the MMSE (Spearman's $r = 0.754$, $P < 0.001$), although the P-SLUMS also exhibited higher sensitivity among low-educated participants (100% vs. the lower sensitivity typically reported for the MMSE). This tool is frequently used to differentiate between cognitive disorders (12). A study reported a sensitivity of 74% for MCI detection, 98% for dementia diagnosis, and 82% for distinguishing dementia from MCI (13).

Methods

This cross-sectional diagnostic accuracy study evaluated the psychometric properties of the Persian version of the SLUMS exam, which was administered to 450 community-dwelling older adults in Tehran. Participants were recruited through convenience sampling; they were referred to one of the Tehran Municipality centers and engaged in a cultural program from March to November 2021. The use of convenience sampling may introduce selection bias and limit the generalizability of our findings.

Inclusion criteria included individuals aged ≥ 60 who could understand and converse in Persian and were willing to participate.

Exclusion criteria included advanced stages of cancer, severe dementia (as indicated by a score greater than 8 on the Persian version of the Global Deterioration Scale (GDS)), severe depression (as indicated by a score greater than 15 on the Persian version of the Patient Health Questionnaire (PHQ-9)), moderate to severe speech disorders, and significant visual impairment.

Data were collected by a trained psychiatrist who conducted interviews with the participants. The questionnaires utilized in this study included a widely used demographic questionnaire, the Persian version of the MMSE (14) to assess concurrent validity, and the Persian-translated version of the SLUMS exam.

a) Questionnaire preparation: The original SLUMS exam was obtained from its official website (15).

b) Translation-back translation: According to the WHO protocol, six steps were followed for the forward–backward translation of the SLUMS exam.

Initially, an expert fluent in Persian translated the tool from English into Persian (Forward Translation). Subsequently, a team comprising one psychiatrist, two psychologists, one epidemiologist, and one language specialist reviewed the forward-translated version and, in collaboration with the translator, made necessary corrections (Review of Forward Translation). Following this, an English-fluent expert, without knowledge of the original tool, back-translated the Persian version into English (Back Translation). The research team then compared the back-translated version with the original SLUMS exam, correcting any inconsistencies or misinterpretations (Comparison of Back-translation with the Original Text). After these reviews, a comprehensive version was produced, encompassing all necessary concepts (Consensus and Finalization). Finally, the finalized version was tested in a pilot study with participants to identify any issues related to comprehension or usability, and feedback was collected to facilitate final adjustments (Pre-testing the Translated Version) (16).

c) Content validity: The content validity of the SLUMS exam was assessed using both qualitative and quantitative methods.

To assess qualitative content validity, the expert panel evaluated each item's grammar, wording, item allocation, and scaling. Based on their feedback, the tool was modified as necessary.

To assess the quantitative content validity, the expert panel calculated the Content Validity Ratio (CVR) and the Content Validity Index (CVI).

The CVR was calculated using a 3-point Likert scale, where 3 indicated "essential," 2 indicated "useful but not essential," and 1 indicated "not essential."

[DOI: 10.47176/mjiri.40.35]

The following formula was utilized: where “N” represents the total number of experts, and “ne” denotes the number of experts who rated each item as “essential.” According to Lawshe’s table, items with a CVR score greater than 0.85 were accepted.

$$\mathbf{CVR}=\frac{ne-\frac{N}{2}}{\frac{N}{2}}$$

The CVI was calculated by the expert panels using the recommendations of Waltz and Bausell.

The items were evaluated based on "simplicity," "relevancy," and "clarity" using a 4-point Likert scale: 1 = not relevant, 2 = somewhat relevant, 3 = relevant, and 4 = totally relevant. Items rated with scores of 3 or 4 were accepted. A CVI value greater than 0.79 was deemed acceptable (17).

d) Face validity: For the qualitative face validity evaluation of the P-SLUMS exam, the expert panel (the same group that evaluated the CVI and CVR) assessed each item based on its "difficulty," "irrelevance," and "ambiguity." Their recommendations were subsequently applied to the items.

For the quantitative evaluation of face validity, the Item Impact score was calculated using a 5-point Likert scale, where 5 indicates "very important," 4 "important," 3 "somewhat important," 2 "slightly important," and 1 "not important." This form of validity was employed to reduce and eliminate inappropriate items. The Impact Score is derived from the product of Frequency (%) and Importance. Items with an Impact Score of ≥ 1.5 were deemed satisfactory (17).

e) Concurrent (criterion) validity: The P-SLUMS exam was compared with the Persian version of the MMSE. A Spearman’s correlation value of ≥ 0.7 was deemed acceptable for evaluating concurrent validity (18).

The sensitivity and specificity of the P-SLUMS exam were determined based on the established cut-off points of the Persian version of the MMSE.

f) Construct validity: Construct validity evaluates the extent to which a tool accurately measures the concept it is intended to assess (19). A total of 450 participants completed the P-SLUMS exam to validate its factor structure. The adequacy of our sample size for factor analysis was determined based on the standard recommendation of including 5–10 participants per item. With 11 items in the P-SLUMS exam, a minimum of 55–110 participants would have been sufficient. Our study, however, included 450 participants, which exceeds this recommendation.

The goodness-of-fit indices for the model, including the relative chi-square ratio (x2/df), Comparative Fit Index (CFI), Parsimonious Normed Fit Index (PNFI), Parsimonious Comparative Fit Index (PCFI), and Root Mean Square Error of Approximation (RMSEA), were calculated. A relative chi-square ratio greater than 3 indicated an acceptable model fit. The CFI ranges from 0 to 1, with values closer to 1 signifying a good model fit. An RMSEA value below 0.08 was considered appropriate. PCFI and PNFI values exceeding 0.5 indicated acceptable model parsimony (19).

g) Reliability:

Stability: The stability of the P-SLUMS exam was assessed using test-retest reliability based on the Intraclass Correlation Coefficient (ICC), a total score of ≥ 0.75 considered indicative of good reliability. Several studies have indicated that a sample size of 15–30 participants is generally adequate for test–retest reliability when the instrument is not expected to exhibit substantial random variability. For instance, Polit and colleagues recommend that 15–30 participants are sufficient for estimating ICC in pilot or validation studies. Furthermore, the COSMIN guidelines deem a minimum of 20 participants acceptable for reliability testing (20-22).

In this context, a sub-sample of 20 participants completed the tool on two occasions, separated by a two-week interval, to evaluate test-retest reliability.

Internal consistency: Internal consistency was evaluated using Cronbach’s alpha, and values above 0.7 were considered acceptable, indicating adequate reliability.

Results

Translation-back translation

According to the WHO protocol for translation and back translation of the tool, six steps were implemented: forward translation, review of the forward translation, back translation, comparison of the back translation with the original text, consensus, finalization, and pre-testing of the translated version.

Data collection

Data were collected from 450 community-dwelling older adults. The mean age of the participants was 69.26 ± 8.82 years, with 274 (60.9%) identifying as female. Among the participants, 266 (59.1%) were married, while 56 (12.5%) lived alone. The majority of participants (233 individuals) had attained a high school level of education (51.8%) (Table 1).

Validity

a) Content validity: An expert panel, consisting of a neurologist, a psychiatrist, a gerontologist, an epidemiologist, two psychologists, and two geriatric nurses, evaluated the content validity of the P-SLUMS exam. In our study, the results of the CVR and CVI indicated that most items demonstrated acceptable content validity. The CVR values ranged from 0.75 to 1.00, meeting Lawshe’s minimum recommendation for eight experts, while the I-CVI values ranged from 0.83 to 1.00, exceeding Lynn’s suggested cut-off of 0.78. On average, the CVR was 0.87, and the I-CVI was 0.94. Overall, these findings indicate that the experts concurred that the items were relevant and clear.

b) Face validity: To evaluate qualitative face validity, the expert panel assessed the face validity and recommended no changes to the items. To determine quantitative face validity, the expert panel rated the importance of all items, with the Impact Score for each item exceeding 1.5; thus, face validity was deemed acceptable. All items of the tool demonstrated an Impact Score greater than 1.5, indicating their appropriateness and relevance for measur-

[DOI: 10.47176/mjiri.40.35]

Table 1. Demographic Characteristics of the Participants (n = 450)

Variables		Number (%)
Gender	Female	274 (60.9%)
	Male	176 (39.1%)
Marital Status	Single	38 (8.4%)
	Married	266 (59.1%)
	Divorced or separated	30 (6.7%)
	Widow	116 (25.8%)
Education Levels	Primary school	76 (16.9%)
	High school	233 (51.8%)
	Diploma	83 (18.4%)
	Academic	58 (12.9%)
Living Arrangement Type	Alone	56 (12.5%)
	With Family	304 (67.6%)
	With Caregiver	90 (19.9%)
Occupation	Employed	93 (20.6%)
	Retired	175 (38.9%)
	Housekeeper	107 (23.8%)
	Unemployed	75 (16.7%)

Table 2. Correlations (Concurrent Validity) Between Sub-scores of the P-SLUMS Exam and the Persian Version of the MMSE

Domain	Spearman's correlation (95% Confidence Interval)	*P*-value
Visual-Spatial and Executive Function	1	< 0.001
Memory: Immediate Recall with Interference	0.620	< 0.001
Attention and Orientation	0.251	< 0.001
MMSE Total Score	0.754	< 0.001

Table 3. Cut-off Points for the P-SLUMS Exam for Differentiation of Participants with Dementia

Level of Education	Cut-off Point	Sensitivity	Specificity	Correctly Classified	LR+	LR-	AUC (95% CI)
< High school	15	100%	76.81%	78.95%	4.313	0	0.905
≥ High school	18	76.15%	72.81%	75.13%	2.801	0.277	0.848

LR, Likelihood Ratio

ing the intended constructs. The mean Impact Score was 2.12 (SD = 0.24), confirming that the items were regarded as important by the experts.

c) Concurrent (criterion) validity: The concurrent validity between the P-SLUMS exam and the Persian version of the MMSE was evaluated. Spearman's correlation between the total scores of these two tools was 0.754 ($P < 0.001$), indicating a strong correlation. This robust Spearman's correlation between the SLUMS and the MMSE (r = 0.754, $P < 0.001$) represents a significant strength of this study. This finding indicates excellent concurrent criterion validity and demonstrates that the Persian version of the SLUMS performs consistently with an established cognitive screening tool (Table 2).

The highest correlation was found in the "Visual-Spatial and Executive Function" domain (Spearman's Correlation = 1, $P < 0.001$), while the lowest correlation was observed in the "Attention and Orientation" domain (Spearman's Correlation = 0.251, $P < 0.001$) (Table 3). Our findings indicate that the "Attention and Orientation" domain exhibited the lowest correlation with the MMSE (r = 0.251, $P < 0.001$), whereas the "Visual-Spatial and Executive Function" domain demonstrated the highest correlation (r = 1.0, $P < 0.001$) among the other domains. Clinically, the "Attention and Orientation" domains are more sensitive to early fluctuations caused by fatigue, mood, or environmental factors, particularly in community-dwelling older adults. Consequently, variability in these scores is anticipated and does not compromise the overall validity of this tool. The high correlation observed in other domains, such as "Visual-Spatial and Executive Function," supports the ability of the P-SLUMS to reliably detect core deficits in cognitive performance, which are most pertinent for the early detection of mild cognitive impairment and dementia (Table 4).

Sensitivity and Specificity

Receiver Operating Characteristic (ROC) curve analysis was performed to assess the efficacy of the P-SLUMS exam in comparison to the Persian version of the MMSE for diagnosing dementia among participants with a high school education or higher, as well as those with less than a high school education.

The cut-off score for the P-SLUMS exam is 18 for the detection of MCI and 19 for the diagnosis of dementia, with a sensitivity of 0.74 and a specificity of 0.82, respectively, utilizing the Persian version of the MMSE's cut-off points of < 23. The Area Under the Curve (AUC) was also evaluated (Figure 1A and Figure 1B). The AUC for dementia detection among participants with a high school education or higher was 0.848, whereas for those with less than a high school education, it was 0.905. These values indicate that the P-SLUMS demonstrates good discriminatory ability across different educational levels.

The cut-off points of the P-SLUMS exam for differentiating participants with dementia were established at 15 for low-educated samples (with a sensitivity of 100% and specificity of 76.15%) and 18 for high-educated samples (with a sensitivity of 76.81% and specificity of 72.81%). This indicates the appropriate sensitivity and specificity of

[DOI: 10.47176/mjiri.40.35]

Table 4. Domains of the P-SLUMS

The SLUMS exam sub-scores		Domains			
			1	2	3
SLUMS 10A	Visual-Spatial and Executive Function	1 Visual-Spatial and Executive Function	0.738		
SLUMS 10B			0.735		
SLUMS 11C	Executive Function, Extrapolation		0.619		
SLUMS 11A			0.646		
SLUMS 7	Delayed Recall with Interference		0.568		
SLUMS 11D	Executive Function, Extrapolation		0.529		
SLUMS 11B			0.499		
SLUMS 9B	Visual-Spatial	2 Memory: Immediate Recall with Interference		0.886	
SLUMS 9A				0.881	
SLUMS 8	Registration, Digit Span			0.602	
SLUMS 6	Memory: Immediate Recall with Interference (time constraint)			0.539	
SLUMS 1	Attention, Immediate Recall, Orientation	3 Attention and Orientation			0.554
SLUMS 2					0.488
SLUMS 3					0.352
SLUMS 5A	Numeric Calculation, Registration				0.829
SLUMS 5B					0.783

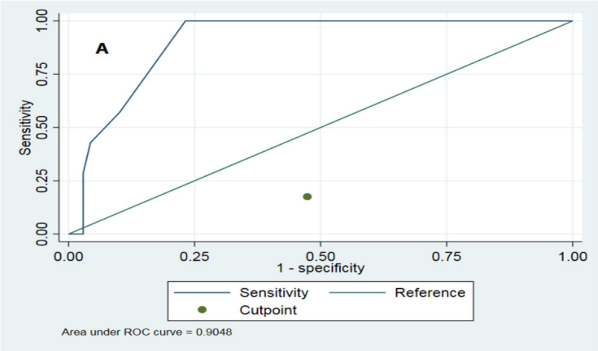

Figure 1A. The ROC Curve Analysis of the P-SLUMS Exam Compared with the Persian Version of the MMSE Cutoff Points for Dementia Diagnosis in Participants with High School Education or Higher

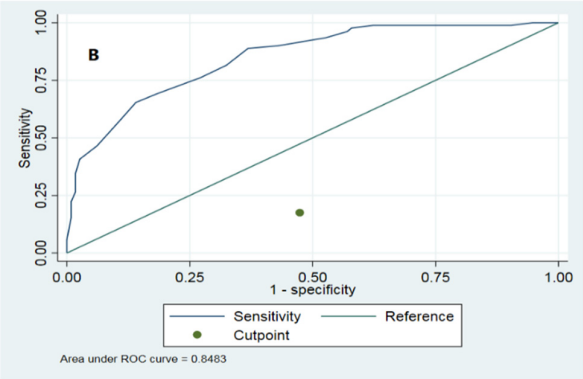

Figure 1B. The ROC curve analysis of the P-SLUMS exam compared with the Persian version of the MMSE cutoff points for dementia diagnosis in participants with less than a high school education

the P-SLUMS exam. Notably, one of the key findings of our study is the exceptional sensitivity of the P-SLUMS exam among participants with lower education, where the sensitivity for dementia detection was 100% (with a specificity of 76.15%). This underscores that the P-SLUMS is particularly suitable for populations with limited education, addressing a significant limitation of other cognitive screening tools, such as the MMSE, which may underestimate impairment in individuals with lower educational attainment.

[DOI: 10.47176/mjiri.40.35]

Confirmatory Factor Analysis (CFA)

In our CFA analysis, the CFI, PCFI, and PNFI values were 0.82, 0.66, and 0.44, respectively. While these values are moderate and fall below the recommended thresholds for excellent parsimony, they are considered acceptable for our sample, which includes individuals with diverse educational backgrounds and a moderate number of items [21] in our tool. The moderate parsimony values may reflect the heterogeneity of cognitive performance within this population. Notably, other fit indices, including CFI (0.82) and relative chi-square (18.37), supported the overall adequacy of the model. Additionally, the RMSEA value was 0.197, and the relative chi-square value of 18.37 ($P < 0.001$) further confirmed the model's adequacy. In this regard, the RMSEA is considered 'marginal', while the recommended threshold for a good fit is below 0.08. This result may reflect the complexity and heterogeneity of cognitive abilities in our sample, as well as the inclusion of participants with varying educational levels. It is important to note that RMSEA is sensitive to model complexity and sample characteristics (Figure 2).

Reliability

The P-SLUMS exam demonstrated excellent internal consistency (Cronbach's alpha = 0.866). The Intraclass Correlation Coefficient (ICC) was 0.979 (95% CI: 0.915-0.995), indicating outstanding stability. The inter-rater reliability of the SLUMS exam domains is presented in Table 5.

Discussion

The original English version of the SLUMS exam serves as a screening tool for the early detection of cognitive impairments, particularly MCI, which may be overlooked by the MMSE. This tool was developed by Tariq et al. using a sample of older individuals from the United States (11). Given its characteristics, the present study aimed to validate the Persian version of the Saint Louis University Mental Status (SLUMS) exam in a cohort of community-dwelling older adults from Tehran. The findings of this study indicate that the Persian version of this tool (P-SLUMS) is effective for detecting cognitive impairments in the early stages among participants, contributing to the existing literature on the SLUMS exam. This aligns with findings from other studies, such as that by Liu et al., which emphasized the necessity for cognitive screening tools to be culturally adapted in order to develop tailored questions and provide a more accurate evaluation of the cognitive status of the target group. Additionally, the study by Spencer et al. demonstrated, following their systematic review of the psychometric properties of the SLUMS exam in various languages, the importance of evaluating the psychometric properties of this tool across diverse populations and cultures (23). Furthermore, the SLUMS exam has been validated among diverse populations, including Chinese, Korean, Spanish, and Turkish individuals. The psychometric properties of the P-SLUMS exam are comparable to those in other languages, reinforcing the SLUMS exam's capability to detect cognitive impairments across different cultures and languages.

Future studies could consider model modifications, such

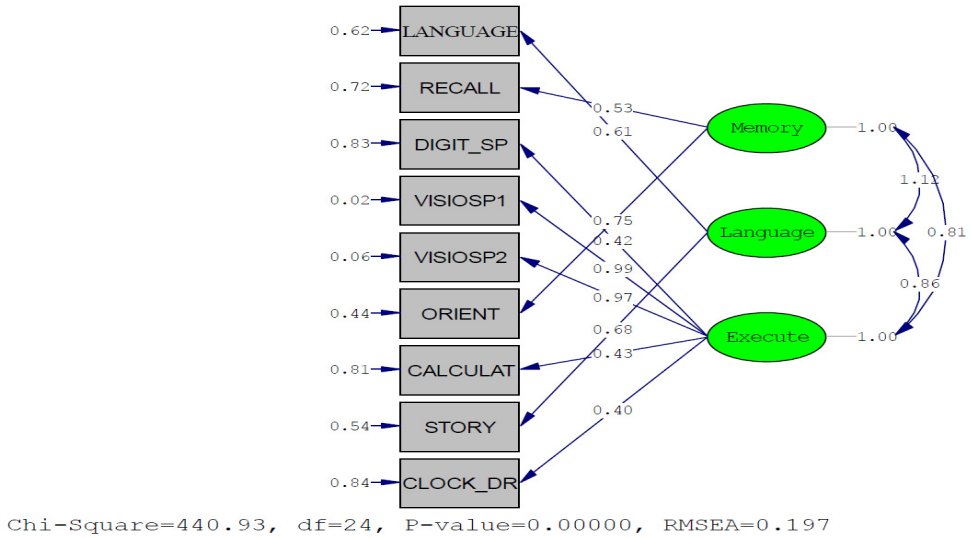

Figure 2. The Results of Confirmatory Factor Analysis of the Persian Version of the Saint Louis University Mental Status Exam

Table 5. Inter-Rater Reliability of the P-SLUMS Exam Domains

P-SLUMS Exam Domains	Number of Domains	ICC (95% Confidence Interval)
Visual-Spatial and Executive Function	SLUMS 1,2,3	1.000
Memory: Immediate Recall with Interference	SLUMS 9,10	1.000
Attention and Orientation	SLUMS 5,6,8	0.947 (0.792-0.987)
Total	All Domains	0.979 (0.915-0.995)

[DOI: 10.47176/mjiri.40.35]

as permitting correlated errors or exploring additional latent factors, to enhance model fit. We have incorporated this discussion into the revised manuscript to acknowledge potential improvements in the factor structure.

Validity

a) Content validity: The CVR and CVI demonstrated the relevance and clarity of the items. These findings are consistent with the study by Tariq et al., which reported strong content validity for the SLUMS exam, indicating that the test content effectively reflects the measurement objectives. However, no specific numerical values were published regarding this validity (11).

b) Face validity: Face validity in other languages has demonstrated comparable results with the P-SLUMS exam, indicating that the tool is suitable for measuring the intended characteristics (23-29).

The high RMSEA (0.197) may indicate the complexity and heterogeneity of cognitive constructs in older adults, as well as the limited number of items per domain. Future studies should consider model modifications, such as allowing correlated errors or exploring additional latent factors, to enhance model fit.

c) Concurrent (criterion) validity: The concurrent validity between the P-SLUMS exam and the Persian version of the MMSE was strong, with a Spearman's correlation of 0.754 ($P < 0.001$), indicating a robust correlation between these two tools and their efficacy in detecting cognitive impairments. Cao et al. also reported a Spearman's correlation of 0.747 between the SLUMS exam and the MMSE scores ($P < 0.001$) (28). Similarly, Zhang et al. found a Spearman's correlation of 0.747 between the SLUMS exam and the MMSE (24). This finding is consistent with previous studies by Tariq et al. and Diniz et al., which demonstrated comparable results between the SLUMS exam and the MMSE when screening for cognitive impairments, particularly in the "Orientation" and "Memory" domains, with reported correlation coefficients ranging from 0.6 to 0.7 (11, 30). Additionally, the study by Vidal et al. demonstrated that the Spanish version of the SLUMS exam exhibited a strong correlation with the MMSE, with a reported correlation coefficient of 0.82 (27).

Confirmatory Factor Analysis (CFA)

The confirmatory factor analysis (CFA) conducted in this study revealed acceptable goodness-of-fit indices for the model, thereby indicating the structural integrity of the P-SLUMS exam. These findings are consistent with the results of studies by Cumming-Vaughn et al. and Rodríguez-Gómez et al., which also demonstrated significant construct validity of the SLUMS exam across various healthcare settings (12, 31).

The CFA confirmed the adequacy of the sample size and the validity of the factor structure of the P-SLUMS exam. Fit indices, including CFI, PCFI, PNFI, and RMSEA, indicated an acceptable model fit, thereby affirming the model's adequacy. These findings are consistent with the results of previous studies, including those by Cumming-Vaughn et al., Zhang et al., Vidal et al., and Rodriguez-Gomez et al., which demonstrated the cross-cultural and linguistic consistency of the SLUMS exam (12, 24, 27, 31).

Sensitivity and Specificity

The diagnostic accuracy of the P-SLUMS exam (Table 3) was comparable to the findings of the study by Cummings-Vaughn et al., which demonstrated an AUC of 0.74 for MCI and 0.98 for dementia (12). Additionally, Lin et al. reported an AUC of 0.88 for the SLUMS exam in the Taiwanese population, underscoring its effectiveness in detecting MCI and more advanced cognitive impairments (32).

The study by Xie et al. demonstrated a superior balance between sensitivity and specificity of the SLUMS exam compared to the MMSE, underscoring its utility as an early diagnostic tool (33). Furthermore, Fong et al. and Tsoi et al. suggested that the SLUMS exam may offer a more comprehensive assessment of cognitive domains, particularly in "Orientation" and "Memory" (34, 35).

Reliability

The P-SLUMS exam demonstrated excellent internal consistency, indicating its reliability within this version of the SLUMS exam. Furthermore, its Intraclass Correlation Coefficient (ICC) was 0.979 (95% CI: 0.915-0.995), reflecting the exceptional stability and test-retest reliability of the P-SLUMS exam over time. This ICC signifies the strong inter-rater reliability of its domains (Table 5), which aligns with findings from other studies, such as the research conducted by Zhang et al. on 468 Chinese-aged individuals, which reported a Cronbach's alpha of 0.723 for the Chinese version of the SLUMS exam, indicating good internal consistency. Additionally, the high ICC of 0.968 for the Chinese version underscored its excellent test-retest reliability. These findings suggest that the P-SLUMS exam is as reliable as the Chinese version of the SLUMS exam (24).

Kaya et al. conducted a study involving 314 Turkish individuals referred to various clinics, demonstrating a Cronbach's alpha of 0.790 for the Turkish version of the SLUMS exam. The ICC for this version was 0.940, indicating high reliability and applicability in clinical settings (25). These findings are comparable to the results of the current study, and the slightly lower Cronbach's alpha observed can be attributed to cultural differences in the assessment of cognitive domains.

The study conducted by Kim et al. involving 280 elderly Koreans demonstrated comparable findings with the P-SLUMS exam in the Korean version of the SLUMS exam, exhibiting a Cronbach's alpha of 0.765 and an ICC of 0.925. These results indicate good psychometric properties, making the tool suitable for diagnosing cognitive impairment in the Korean elderly population (26). Although these values are slightly lower than those of the original SLUMS exam and the P-SLUMS exam, they still reflect the strong reliability of this instrument. The observed variations may be attributed to differences in sample size or the cognitive profiles of the participants. The Spanish version of the SLUMS exam was validated by

[DOI: 10.47176/mjiri.40.35]

Vidal et al. on 342 Spanish-speaking elderly individuals in 2022, yielding a Cronbach's alpha of 0.743 and an ICC of 0.935. These results underscore the utility of this tool for screening neurocognitive impairments among elderly Spanish speakers, particularly for the early detection of MCI and dementia. Although the ICC for the Spanish version was slightly lower than that of the Persian version, it still demonstrated excellent temporal stability (27). The marginally higher Cronbach's alpha of the P-SLUMS exam compared to the Spanish version may be attributed to the precise translation of the tool, specific characteristics of the sample, or cultural and linguistic differences.

Cao et al. reported similarly high ICC values for the Chinese version of the SLUMS exam, demonstrating the tool's reliability across various linguistic contexts (28). This study indicates a higher diagnostic accuracy for MCI and dementia at early stages compared to the Persian version of the MMSE, with enhanced predictive validity (28).

The strong ICC observed in this study underscores the stability of the SLUMS exam. Comparable studies examining other language versions of the SLUMS exam, such as the research conducted by Noyes et al., reported acceptable internal consistency (Cronbach's alpha of 0.709) and temporal stability (ICC = 0.723) in a veteran sample, thereby reinforcing the conclusion that the SLUMS exam possesses sufficient reliability (29).

A Spearman correlation of 1.0 between domains is highly unlikely in real data. This unusual finding may be due to, for instance, overlapping item content across domains, small sample size, high multicollinearity. Additionally, a Spearman's correlation (100% sensitivity and moderate specificity) may suggest cutoff selection bias, and that MMSE is not a gold standard, as well as, the perfect sensitivity (100%) combined with moderate specificity may reflect cutoff selection bias. Consequently, the reported accuracy may overestimate the true diagnostic performance.

In summary, a comparison of the various types of validity (Content Validity, Face Validity, Concurrent Validity, and Construct Validity) of the Persian version of the SLUMS exam with the original SLUMS exam and its other linguistic versions supports the validity of the P-SLUMS exam. Furthermore, the analysis of the reliability indices indicated a higher value of internal consistency. The ICC values, which reflect test-retest reliability, were relatively consistent across all five language versions. The Persian version demonstrated one of the highest ICC values, compared to the Chinese (0.968), Spanish (0.935), Turkish (0.940), and Korean (0.925) versions. These findings indicate that the SLUMS exam is a valid and reliable tool for screening cognitive impairments. Additionally, the psychometric properties of the P-SLUMS exam are comparable to those of other language versions, demonstrating the usefulness of the SLUMS exam across diverse cultures and populations. These results confirm that the SLUMS exam, irrespective of the language, possesses acceptable internal consistency and test-retest reliability, making it an appropriate tool for screening neurocognitive impairments across various cultures and populations.

Ethics

The study was approved by the Ethics Committee of Endocrinology and Metabolism Research Institute (Ethical Code: EC-00382). All participants provided written informed consent. Regarding cognitive impairment, written informed consent was obtained from a close relative or legal guardian.

Strengths

The primary strength of this study lies in its large sample size, which substantiates the reliability of the P-SLUMS exam findings. Additionally, a notable strength of this study is the meticulous translation of the P-SLUMS. Furthermore, an important aspect of this study is the validation of the P-SLUMS exam, which enriches the existing literature and illustrates the adaptability of this tool across diverse languages and cultures.

Limitations

While the findings of this study are valuable, six limitations should be acknowledged and addressed in future research. First, the sample was limited to a group of community-dwelling older adults from Tehran, which restricts the generalizability of the findings to other regions of Tehran and Iran. Second, the use of convenience sampling presents a limitation that may introduce selection bias, and may lead to limited external validity; and avoid broad population-level conclusions.

Additionally, cultural adaptations beyond mere translation (e.g., cognitive styles) were not specifically tested. Third, although the psychometric properties of the P-SLUMS exam were deemed acceptable in this study, its performance in clinical settings, such as hospitals or long-term care facilities, has not been evaluated. Fourth, to ascertain whether the P-SLUMS exam can accurately detect cognitive changes, future studies should focus on the predictive validity of the P-SLUMS exam, as MCI is a progressive condition; our study—being cross-sectional—was unable to examine whether individuals with MCI eventually transition to dementia. Assessing the predictive ability of the P-SLUMS requires longitudinal follow-up, which was beyond the scope of the present work. Fifth, the relatively high RMSEA observed in our analysis may be considered a limitation. Future studies with larger samples or alternative factor structures might yield a better model fit. Sixth, although the Persian version of SLUMS was carefully translated and underwent content validation, we did not evaluate deeper cultural factors such as cognitive styles, typical problem-solving approaches, or culturally shaped memory habits. These elements may also influence performance and should be explored in future research.

Conclusion

The Persian version of the SLUMS exam is a valid and reliable instrument for the early detection of cognitive impairments, including MCI and dementia among participants. Notably, the P-SLUMS exam exhibits appropriate psychometric properties.

[DOI: 10.47176/mjiri.40.35]

Acknowledgment

The authors would like to acknowledge the Endocrinology and Metabolism Research Institute of Tehran University of Medical Sciences for their financial support in conducting this study.

Conflict of Interests

The authors declare that they have no competing interests.

Authors' Contributions

F.S., H.F., S.S., M.G., and M.A. conceived and designed the study. F.S. conducted data analysis and interpretation. S.K.S., Z.A.M.K., N.M., and F.G. searched for the needed references and drafted the preliminary version of the manuscript. S.N. wrote and finalized the manuscript. F.S., M.G., and M.A. supervised the study and reviewed the final manuscript. All authors reviewed and approved the final version of the manuscript. M.G. and M.A. are corresponding authors and take responsibility for the overall integrity of the work.

Ethical Considerations

This study was approved by the Ethics Committee of the Endocrinology and Metabolism Research Institute of Tehran University of Medical Sciences (approval code: EC-00382).

Funding Support

The Elderly Health Research Center of the Endocrinology and Metabolism Research Institute of Tehran University of Medical Sciences provided financial support for this study.

Data availability

All data generated or analyzed during this study are included in this published article.

AI Use Statement

N/A.

References

1. Béjot Y, Yaffe K. Ageing Population: A Neurological Challenge. Neuroepidemiology. 2019;52(1-2):76-8.
2. Shepherd H, Livingston G, Chan J, Sommerlad A. Hospitalisation rates and predictors in people with dementia: a systematic review and meta-analysis. BMC Med. 2019 Jul 15;17(1):130. doi: 10.1186/s12916-019-1369-7. PMID: 31303173; PMCID: PMC6628507.
3. Yliranta A, Nuorva J, Karjalainen V-L, Ahmasalo R, Jehkonen M. The dementia apraxia test can detect early-onset Alzheimer's disease. Neuropsychology. 2023;37(1):44.
4. Harvey PD. Domains of cognition and their assessment. Dialogues Clin Neurosci. 2019;21(3):227-37.
5. Freitas S, Simões MR, Marôco J, Alves L, Santana I. Construct validity of the Montreal Cognitive Assessment (MoCA). J Int Neuropsychol Soc. 2012;18(2):242-50.
6. Wu Y, Verdú S. MMSE domain. IEEE Trans Inf Theory. 2011;57(8):4857-79.
7. American Speech-Language-Hearing Association. Pros and cons of screening tools for assessing dementia [Internet]. Available from: https://leader.pubs.asha.org/do/10.1044/pros-and-cons-of-screening-tools-for-assessing-dementia/full/
8. Saczynski JS, Inouye SK, Guess J, Jones RN, Fong TG, Nemeth E, Hodara A, Ngo L, Marcantonio ER. The Montreal Cognitive Assessment: Creating a Crosswalk with the Mini-Mental State Examination. J Am Geriatr Soc. 2015 Nov;63(11):2370-4. doi: 10.1111/jgs.13710. Epub 2015 Oct 27. PMID: 26503296; PMCID: PMC4704083.
9. Shwartz SK, Morris RD, Penna S. Psychometric properties of the Saint Louis University Mental Status Examination. Appl Neuropsychol Adult. 2019;26(2):101-10.
10. Stern S. Psychometric Properties of the Saint Louis University Mental Status Examination (SLUMS) for the Identification of Mild Cognitive Impairment (MCI) in a Veteran Sample. Dissertation, Georgia State University; 2014. doi: https://doi.org/10.57709/5272343
11. Tariq SH, Tumosa N, Chibnall JT, Perry MH, Morley JE. Comparison of the Saint Louis University Mental Status Examination and the Mini-Mental State Examination for detecting dementia and mild neurocognitive disorder—A pilot study. Am J Geriatr Psychiatry. 2006;14(11):900-10. doi: 10.1097/01.JGP.0000221510.33817.86.
12. Cummings-Vaughn LA, Chavakula NN, Malmstrom TK, Tumosa N, Morley JE, Cruz-Oliver DM. Veterans Affairs Saint Louis University Mental Status examination compared with the Montreal Cognitive Assessment and the Short Test of Mental Status. J Am Geriatr Soc. 2014 Jul;62(7):1341-6. doi: 10.1111/jgs.12874. Epub 2014 Jun 10. PMID: 24916485.
13. Cruz-Oliver DM, Malmstrom TK, Roegner M, Tumosa N, Grossberg GT. Cognitive deficit reversal as shown by changes in the Veterans Affairs Saint Louis University Mental Status (SLUMS) examination scores 7.5 years later. J Am Med Dir Assoc. 2014 Sep;15(9):687.e5-10. doi: 10.1016/j.jamda.2014.05.004. Epub 2014 Jun 20. PMID: 24953686.
14. Ansari NN, Naghdi S, Hasson S, Valizadeh L, Jalaie S. Validation of a Mini-Mental State Examination (MMSE) for the Persian population: a pilot study. Appl Neuropsychol. 2010 Jul;17(3):190-5. doi: 10.1080/09084282.2010.499773. Erratum in: Appl Neuropsychol. 2010 Oct;17(4):312. PMID: 20799110.
15. Rehabilitation Institute of Chicago. Saint Louis University Mental Status [Internet]. Available from: https://www.sralab.org/rehabilitation-measures/saint-louis-university-mental-status-
16. World Health Organization. Process of translation and adaptation of instruments [Internet]. Available from: https://www.who.int/substance_abuse/research_tools/translation/en/
17. Fitzner K. Reliability and validity: a quick review. Diabetes Educ. 2007;33:775-80.
18. Scribbr. Concurrent validity [Internet]. Available from: https://www.scribbr.com/methodology/concurrent-validity/
19. Scribbr. Construct validity [Internet]. Available from: https://www.scribbr.com/methodology/construct-validity/
20. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007;60(1):34–42.
21. Koo TK, Li MY. A guideline for selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 2016;15(2):155–163.
22. Sousa VD, Rojjanasrirat W. Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: a clear and user-friendly guideline. J Eval Clin Pract. 2011;17(2):268–274.
23. Spencer RJ, Noyes ET, Bair JL, Ransom MT. Systematic Review of the Psychometric Properties of the Saint Louis University Mental Status (SLUMS) Examination. Clin Gerontol. 2022 May-Jun;45(3):454-66. doi: 10.1080/07317115.2022.2032523. Epub 2022 Feb 2. PMID: 35107414.
24. Zhang S, Wu Y-H, Zhang Y, Zhang Y, Cheng Y. Preliminary study of the validity and reliability of the Chinese version of the Saint Louis University Mental Status Examination (SLUMS) in detecting cognitive impairment in patients with traumatic brain injury. Appl Neuropsychol Adult. 2021;28(6):633-40.
25. Kaya D, Karaca F, Bozkurt M. Turkish validation and reliability of the Saint Louis University Mental Status Examination. Turk J Geriatr. 2020;23(3):212-20. doi: 10.4274/tjg.2020.8934.
26. Kim SY, Park E, Choi H. Validity and reliability of the Korean version of the Saint Louis University Mental Status (SLUMS-K) examination for screening of dementia. J Korean Geriatr Psychiatry. 2021;25(3):177-85. doi: 10.3390/jkgp20210723.

[DOI: 10.47176/mjiri.40.35]

27. Vidal S, Aguilar-Navarro S, Zamora-García M, Guzmán L. Psychometric evaluation of the SLUMS screening tool in a Spanish-speaking elderly population. Geriatr Gerontol Int. 2022;22(4):329-36. doi: 10.1111/ggi.14426.
28. Cao L, Hai S, Lin X, Shu D, Wang S, Yue J, et al. Comparison of the Saint Louis University Mental Status Examination, the Mini-Mental State Examination, and the Montreal Cognitive Assessment in detection of cognitive impairment in Chinese elderly from the geriatric department. J Am Med Dir Assoc. 2012 Sep;13(7):626-9. doi: 10.1016/j.jamda.2012.05.003.
29. Noyes ET, Major S, Wilson AM, Campbell EB, Ratcliffe LN, Spencer RJ. Reliability and Factor Structure of the Saint Louis University Mental Status (SLUMS) Examination. Clin Gerontol. 2022;1-7.
30. Diniz BS, Yassuda MS, Nunes PV, Radanovic M, Forlenza OV. Mini-Mental State Examination performance in mild cognitive impairment subtypes. Int Psychogeriatr. 2013;25(4):689-97. doi: 10.1017/S1041610212001954.
31. Rodríguez-Gómez O, Sanabria Á, Esteban E, Gil P, Peña-Casanova J, Boada M. The Saint Louis University Mental Status (SLUMS) examination for screening mild cognitive impairment and dementia in a Spanish population. Neuropsychiatr Dis Treat. 2017;13:1745-51. doi: 10.2147/NDT.S137759.
32. Lin YJ, Chen YC, Wang CL. A comparison of SLUMS, MMSE, and MoCA for detecting MCI and dementia in Taiwanese populations. J Formos Med Assoc. 2019;118(5):967-74. doi: 10.1016/j.jfma.2018.09.008.
33. Xie J, Brayne C, Matthews FE. Survival times in people with dementia: Analysis from population-based cohort studies with 14 years of follow-up. BMJ. 2021;373:n748. doi: 10.1136/bmj.n748.
34. Fong TG, Davis D, Growdon ME, Albuquerque A, Inouye SK. The interface between delirium and dementia in elderly adults. Lancet Neurol. 2020;18(3):273-9. doi: 10.1016/S1474-4422(19)30360-6.
35. Tsoi KK, Chan JY, Hirai HW, Wong SY, Kwok TC. Cognitive tests to detect dementia: A systematic review and meta-analysis. JAMA Intern Med. 2015;175(9):1450-8. doi: 10.1001/jamainternmed.2015.2152.

[DOI: 10.47176/mjiri.40.35]