Research Article

The Persian Version of the Abbreviated Word Auditory Recognition and Recall Measure: Validity and Reliability Assessment

Romina Shahrokhi[1], Maryam Ramezani[1,2*], Mohanna Javanbakht[1,2], Behrooz Mahmoodi-Bakhtiari[3], Enayatollah Bakhshi[4]

[1.] Department of Audiology, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran
[2.] Pediatric Neurorehabilitation Research Center, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran
[3.] Department of Performing Arts, University of Tehran, Tehran, Iran
[4.] Department of Epidemiology, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran

ORCID ID:
Romina Shahrokhi: 0009-0005-7478-6573
Maryam Ramezani: 0000-0003-1556-6130
Mohanna Javanbakht: 0000-0002-2876-3208
Behrooz Mahmoodi-Bakhtiari: 0000-0002-5671-4040
Enayatollah Bakhshi: 0000-0001-8049-0190

Citation: Shahrokhi R, Ramezani M, Javanbakht M, Mahmoodi-Bakhtiari B, Bakhshi E. The Persian Version of the Abbreviated Word Auditory Recognition and Recall Measure: Validity and Reliability Assessment. Aud Vestib Res. 2025;34(3):?-?.

Article info:
Received: 02 Sep 2024
Revised: 19 Nov 2024
Accepted: 20 Nov 2024

[*] **Corresponding Author:** Department of Audiology, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran. maryamramezani16@gmail.com

Short running title: The Persian Version of the Abbreviated Word…

Highlights:

- The Persian WARRM (P WARRM) has high validity and reliability
- The P WARRM was validated to measure the auditory WM of Persian speaker
- The P WARRM score strongly correlates with forward/backward digit span tests

ABSTRACT

Background and Aim: Speech perception in noise involves recognizing speech sounds. Given the importance of working memory in speech perception and the lack of an auditory Working Memory (WM) test in Persian with appropriate lexical load, this study aimed to validate the Persian version of the Word Auditory Recognition and Recall Measure (WARRM) in normal-hearing Persian speakers.

Methods: The study involved two main stages: developing the Persian version and evaluating its validity and reliability. Participants included 59 monolinguals, Persian-speaking, normal-hearing college students aged 18–25 in Tehran, Iran, in 2023. Persian words were selected with the help of a linguistics expert based on phonetic and semantic features. Face and content validity were assessed by 10 experts. The WARRM test and the forward

and backward digit span tests were administered. To assess the test-retest reliability, the test was performed by 30 participants at a 2-week interval and measured by the Intraclass Correlation Coefficient (ICC).
Results: The WARRM test showed good face and content validity. The internal consistency was 0.72 for the overall test, 0.49 for the recognition subtest, 0.73 for the judgment subtest, and 0.87 for the recall subtest. The ICC values were 0.88 (overall), 0.89 (recall), and 0.48 (recognition). Significant positive correlations were found between the WARRM score and the scores of the forward digit span ($r=0.65$, $p<0.001$) and backward digit span ($r=0.43$, $p<0.001$).
Conclusion: The Persian version of the abbreviated WARRM has good validity and reliability for use in clinical and rehabilitation studies on Persian speaker.
Keywords: Auditory working memory; recognition; recall; validity; reliability

Introduction

Speech perception in noise is not limited to recognizing the sequence of speech sounds, even when the listener's task is merely to repeat the given sentence [1]. It depends on the interaction of sensory and cognitive processes, involving both bottom-up and top-down processes [2]. Cognition refers to a wide range of activities in the human brain for acquiring knowledge and understanding, including perceiving, thinking, reasoning, memory, analyzing, planning, attention, creativity, judgment, awareness, and insight [3]. Although various cognitive abilities are involved in speech comprehension, more attention has been paid to the effects of Working Memory (WM) [4]. WM is the ability to actively maintain and manipulate information in the mind for performing complex tasks such as reasoning, perception, and learning [5]. WM is crucial in speech perception, especially in complex auditory situations.

There is a need to assess auditory WM in the clinical evaluation of hearing [6]. Its evaluation can be helpful for diagnostic and auditory rehabilitation purposes. Various tests exist to assess auditory working memory using different materials (words, sentences, numbers), different stimuli (visual, auditory), and different response methods (speaking, ordering, pointing). Some WM tests have been validated in Iran, including the N-back test [7, 8], reading span test [9], digit span test (forward and backward), and non-word repetition test [10]. However, some studies have shown that these tests do not correlate well with the ability to understand speech in noise [11, 12], which may be due to inappropriate test materials [7, 12]. Although Aghamolaei et al. developed and evaluated the Persian version of the Dichotic Auditory Verbal Memory Test (DAVMT) to assess auditory-verbal memory, the tool is a dichotic test and does not exclusively assess memory, potentially involving dichotic disorders and corpus callosum issues [13]. The Word Auditory Recognition and Recall Measure (WARRM) is a relatively new test designed to simultaneously assess word recognition and auditory WM performance using word stimuli in adults [14]. This test also has an abbreviated version developed by Smith et al. in 2020 [15]. Their study indicated that both original and abbreviated versions of WARRM had recall and recognition scores correlated with other memory tests.

The WARRM may provide a more practical and relevant measure compared to other auditory WM tests. It has a short duration. On the other hand, the use of speech stimuli for evaluating the interaction between auditory and cognitive processing experienced in everyday listening is more valid [16]. Considering the importance of WM in speech perception and given that previous studies have indicated that lexical load in test materials can affect outcomes [17], and there is no psychometrics study to validate the WARRM for Persian speaker, this study aimed to develop and assess the validity and reliability of the Persian version of abbreviated WARRM.

Methods

This study was conducted in two stages: development of the Persian version and psychometric evaluation. Participants were 59 normal-hearing, Persian-speaker aged 18–25 from the universities in Tehran, Iran. In the first stage, after obtaining permission from the original version's developer, the Persian version was prepared and analyzed by a linguistic expert for phonetic, semantic, and syllabic features. Many monosyllabic words from everyday language and sources, such as the Academy of Persian Language, were gathered and categorized into consonant-vowel, consonant-vowel-consonant, and consonant-vowel-consonant-consonant groups. Words with multiple meanings, proper nouns (names of places or people), prepositions, and verbs, words borrowed from other languages, words with uncommon or limited usage, and words with different pronunciations or unstable meanings were excluded. A list of 100 balanced monosyllabic words was finally prepared. Achieving homogeneity for test materials in phonetic balance has been a topic of interest among researchers. However, there

is a lack of consensus on the effectiveness of this approach [18-21]. This list was balanced based on the phonetic approach with the same focus on high-frequency and low-frequency sounds in selected vowels and consonants.
In the second stage, for content and face validity evaluation, an email was sent to eight audiologists and two linguistics accompanied by explanations about the test and word selection process. The face validity was confirmed by these 10 specialists and by five people with middle-high school education. For face validity, reviewers were asked to rate each word on a scale from 1 (very weak) to 5 (very strong). For determining the Content Validity Ratio (CVR) based on Lawshe's method, each item was scored as "essential", "useful but not essential", and "not essential". We also assessed the Content Validity Index (CVI) in terms of "relevance" according to Waltz and Bausell's method [22]. For each word, the experts were asked to rate it as "not relevant", "somewhat relevant", "relevant", and "highly relevant". Given the number of audiologists evaluating the validity, an acceptable CVR value is above 0.62, and the acceptable CVI for each item is more than 0.79. After the CVR and CVI evaluations for each word, necessary changes were made based on experts' comments. The final selected words that met all criteria were recorded by a female speaker with a clear and familiar voice in the Golbang Institute in Tehran, Iran. Words with uncommon pronunciation were replaced. The recorded words were categorized into five 20-word lists in the Audacity sound editing software. Similar to the English version of the WARRM, each list in the Persian version had set sizes of 2, 3, 4, 5, and 6 items (20 in total). The participants signed a consent form, and their medical history and educational level were recorded. The results confirmed their normal hearing after performing otoscopy, tympanometry, and audiometry. They should have a score >26 on the Montreal Cognitive Assessment (MoCA). Exclusion criteria were unwillingness to continue participation, poor attention, and lack of cooperation.
The Persian WARRM test was administered using an ASUS laptop (made in Taiwan) and TSCO headphones calibrated with an L-2250 sound level meter (Brüel & Kjær Co., Denmark). Each test list contained 20 words arranged in sets of 2, 3, 4, 5, and 6 words. The WARRM provides two scores: word recognition and recall. The recognition score is the percentage of correctly recognized words, and the recall score is the percentage of correctly recalled words. Participants first heard a carrier phrase, "you will cite" followed by a target word after 50 ms. Participants had 3 seconds to repeat the word before the onset of the next carrier phrase; if unsure, they could guess. After repeating the target word (word-recognition task), participants judged if the first letter of each target word was in the first half or the second half of the alphabet. The judgment score was calculated in terms of overall percent correct but used only to ensure that participants were engaged in the task. The judgment scores should be sufficiently high to have confidence that the listener was engaged in the alphabet task and not just in guessing. This judgment task had no effect on the overall WAARM score. Recognition and judgments had to be done quickly as there was a 3-s interval between the words. After repeating and judging each word, participants recalled all words in sequence after hearing a 500-Hz beep (recall task). The examiner had a scoring sheet to mark and score the correct words. The WARRM materials were presented using a playback device and delivered binaurally through headphones at 60 dB.
We used forward and backward digit span tests to examine the correlation of WARRM test score with the scores of these WM tests. Pearson's correlation coefficient was used to determine their correlation. The Kuder-Richardson formula was used to evaluate the internal consistency of the test since the data were dichotomous. To assess the test-retest reliability, the test was performed by 30 participants at a 2-week interval. The Intraclass Correlation Coefficient (ICC) was used to measure test-retest reliability. Data analysis was conducted in SPSS v.17 software.

Results

Demographic and baseline information of the participants are shown in Table 1. The participants were 59 college students with a mean age of 21.41±2.48 years. They included 62.7% females, and 91.5% were right-handed. For face validity of the test, the average scores given by the judges for all words are above 4 (out of 5), with a mean score of 4.48±0.13, indicating that the Persian version of the WARRM had acceptable face validity. The CVR value for all words was above 0.62, with a mean score of 0.68±0.27. The CVI value was above 0.79 with a mean score of 0.81±0.31. Therefore, the test had a good content validity.
Using the Kuder-Richardson formula, internal consistency for the overall test was 0.72. For the subscales, the intraclass correlation value was 0.49 for the recognition subscale, 0.73 for the judgment subscale, and 0.87 for the recall subscale. Table 2 shows the mean retest scores of the WARRM and its subscales. For the test-retest reliability, the ICC value was 0.88 for judgment, 0.89 for recall, 0.48 for recognition, and 0.92 for the overall test.

Using the Pearson's correlation test, a significant correlation was found between the WARRM score and the scores of forward digit span ($r=0.65$, $p<0.001$) and backward digit span ($r=0.43$, $p=0.001$). This indicates a direct and significant correlation between the scores of the WARRM test and the forward and backward digit span tests.

Discussion

WM refers to a brain system that temporarily stores and manipulates the information necessary for complex cognitive tasks such as language comprehension, learning, and reasoning [23].

The WARRM is a test for assessing auditory WM. Its initial English version was created in 2016 by Smith et al. [14]. The abbreviated version was later developed in 2020 by Smith et al. [15]. The main purpose of this study was to create a set of Persian monosyllabic words for the abbreviated WARRM to measure the auditory WM in Persian speaker. In this regard, five phonetically balanced and equivalent lists of 20 monosyllabic words were developed. The content validity, face validity, test-retest reliability, and internal consistency of the Persian WARRM were evaluated. The results showed that the Persian version of the WARRM had high validity and reliability based on CVR, CVI, and ICC values.

The WARRM scores of the participants in our study (59 normal-hearing college students; mean age: 21.41±2.48 years) were close to those reported in Smith et al.'s study [14]. In their study, the mean scores of recognition and recall were 99 and 86, respectively. Three groups of listeners participated in their study, including 48 younger listeners with normal hearing (pure-tone thresholds ≤20 dB HL for 250–8000 Hz; mean age: 22.8±2.7 years; 17 males and 31 females), 48 older listeners with normal hearing (pure-tone thresholds ≤25 dB HL at 250–3000 Hz; mean age: 66.9±5.1 years; 8 males and 40 females), and 48 older listeners with sensorineural hearing loss. They reported a significant correlation between the WARRM score and the digit span sequencing test score in participants ($r=0.45$, $p<0.001$) [14]. Smith and Pichora-fuller examined the correlation between WARRM and reading span test scores. Participants were a group of younger listeners with normal hearing and a group of older listeners with hearing loss (n=24). They found a significant correlation ($r=0.55$, $p<0.001$) [16]. In our study, there was a direct and significant correlation between the abbreviated WARRM score and the scores of forward and backward digit span tests ($r=0.65$ and 0.43, respectively, $p<0.001$), which is consistent with the mentioned study. Therefore, it seems that the Persian version of the abbreviated WARRM can yield results similar to other tests for auditory WM in the Persian language. Considering the importance of cognitive abilities, including WM spans in speech perception in noise, it is recommended that the correlation of the Persian WARRM test score with the scores of other speech perception in noise tests be evaluated. Also, the use of this test for rehabilitation and monitoring purposes can be examined in future studies.

This study had some limitations, given that the original version of this test has recently been developed, and there are no validated versions in other languages to compare the results. In the WARRM test, multiple auditory and cognitive structures are involved, the roles of which have remained unclear.

Conclusion

The Persian version of the abbreviated word auditory recognition and recall measure is a valid and reliable tool that can be used as a non-invasive clinical and research test for assessing auditory Working Memory (WM) in Persian speaker. Its score correlates directly and significantly with the scores of forward and backward digit span tests. Therefore, it can be used along with other WM tests to measure lexical burdens.

Ethical Considerations

Compliance with ethical guidelines

In this study, the research team has considered and applied ethical guidelines. The Ethics Committee of the University of Social Welfare and Rehabilitation Science approved this study method (Ethical Code: IR.USWR.1400.337).

Funding

The study was extracted from the MSc thesis at the Department of Audiology, the University of Social Welfare and Rehabilitation Science, Tehran, Iran, and also received a research grant (4027) from the Pediatric Neuroscience Research Center at this University.

Authors' contributions

RS: Study design, acquisition of data, interpretation of the results, statistical analysis, and drafting the manuscript; MR and MJ: Study design, interpretation of the results, and drafting the manuscript; BMB: Interpretation of the results and drafting the manuscript; EB: Statistical analysis.

Conflict of interest

The authors declared no conflict of interest.

Acknowledgments

The authors would like to thank all participants for them cooperation.

References

1. Gussekloo J, de Bont LE, von Faber M, Eekhof JA, de Laat JA, Hulshof JH, et al. Auditory rehabilitation of older people from the general population--the Leiden 85-plus study. Br J Gen Pract. 2003;53(492):536-40.
2. Ziegler JC, Pech-Georgel C, George F, Alario FX, Lorenzi C. Deficits in speech perception predict language learning impairment. Proc Natl Acad Sci U S A. 2005;102(39):14110-5. [DOI:10.1073/pnas.0504446102]
3. Lotfi Y, Kargar S, Javanbakht M, Biglarian A. Development, Validity and Reliability of the Persian Version of the Consonant-Vowel in White Noise Test. JRSR. 2016;3(2):29-34. [DOI:10.30476/jrsr.2016.41090]
4. Akeroyd MA. Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. Int J Audiol. 2008;47 Suppl 2:S53-71. [DOI:10.1080/14992020802301142]
5. Knudsen EI. Fundamental components of attention. Annu Rev Neurosci. 2007;30:57-78. [DOI:10.1146/annurev.neuro.30.051606.094256]
6. Banh J, Singh G, Pichora-Fuller MK. Age affects responses on the Speech, Spatial, and Qualities of Hearing Scale (SSQ) by adults with minimal audiometric loss. J Am Acad Audiol. 2012;23(2):81-91; quiz 139-40. [DOI:10.3766/jaaa.23.2.2]
7. Shokuhifar G, Javanbakht M, Vahedi M, Mehrkian S, Aghadoost A. The relationship between speech in noise perception and auditory working memory capacity in monolingual and bilingual adults. Int J Audiol. 2024:1-8. [DOI:10.1080/14992027.2024.2328556]
8. Mehrkian S, Mozaffari Z, Bakhshi E. The relationship between working memory capacity and temporal and dichotic auditory processing in teachers. Aud Vestib Res. 2019;28(2):100-5. [DOI:10.18502/avr.v28i2.864]
9. Najjari R, Mohammadi M. The Development of Reading and Operation Span Tasks in Persian as Measures of Working Memory Capacity for Iranian EFL Learners. J Teach Lang Skills. 2017;36(2):129-62. [DOI:10.22099/JTLS.2017.24688.2215]
10. Nejati V, Alipour F. Persian version of digit span test, word span, non-word span, and evaluate the psychometric properties and comparable sensitivity in measuring working memory of children. J Appl Psychol. 2016;10(2):73-88.
11. Fostick L, Ben-Artzi E, Babkoff H. Aging and speech perception: beyond hearing threshold and cognitive ability. J Basic Clin Physiol Pharmacol. 2013;24(3):175-83. [DOI:10.1515/jbcpp-2013-0048]
12. Sheft S, Shafiro V, Wang E, Barnes LL, Shah RC. Relationship between Auditory and Cognitive Abilities in Older Adults. PLoS One. 2015;10(8):e0134330. [DOI:10.1371/journal.pone.0134330]
13. Aghamollaei M, Tahaei SA, Jafari Z, Toufan R, Keyhani MR. [Development and evaluation of the Persian version of the dichotic auditory-verbal memory test in 18- to 25-year old normal individuals]. Audiol. 2011;20(2):86-94. Persian.
14. Smith SL, Pichora-Fuller MK, Alexander G. Development of the Word Auditory Recognition and Recall Measure: A Working Memory Test for Use in Rehabilitative Audiology. Ear Hear. 2016;37(6):e360-76. [DOI:10.1097/AUD.0000000000000329]
15. Smith SL, Ryan DB, Pichora-Fuller MK. Development of Abbreviated Versions of the Word Auditory Recognition and Recall Measure. Ear Hear. 2020;41(6):1483-91. [DOI:10.1097/AUD.0000000000000869]
16. Smith SL, Pichora-Fuller MK. Associations between speech understanding and auditory and visual tests of verbal working memory: effects of linguistic complexity, task, age, and hearing loss. Front Psychol. 2015;6:1394. [DOI:10.3389/fpsyg.2015.01394]
17. Kowialiewski B, Lemaire B, Portrat S. How does semantic knowledge impact working memory maintenance? Computational and behavioral investigations. J Mem Lang. 2021;117:104208. [DOI:10.1016/j.jml.2020.104208]
18. Jarolahi F, Delphi M, Tahaie SA, Modarresi Y, Kamali M, Jafari M. [Selection of preeminent list in word recognition score test for adult with normal hearing]. J Res Rehabil Sci. 2012;8(2): 212-8. Persian. [DOI:10.22122/JRRS.V8I2.343]
19. Egan JJ. Basic aspects of speech audiometry. Ear Nose Throat J. 1979;58(5):190-3.
20. Martin FN, Champlin CA, Perez DD. The question of phonetic balance in word recognition testing. J Am Acad Audiol. 2000;11(9):489-93; quiz 522. [DOI:10.1055/s-0042-1748141]
21. Harris RW, Nissen SL, Pola MG, McPherson DL, Tavartkiladze GA, Eggett DL. Psychometrically equivalent Russian speech audiometry materials by male and female talkers. Int J Audiol. 2007;46(1):47-66. [DOI:10.1080/14992020601058117]
22. Waltz CF, Bausell RB. Nursing research: Design, statistics, and computer analysis. Philadelphia: FA Davis company; 1981.
23. Baddeley A. Working memory. Science. 1992;255(5044):556-9. [DOI:10.1126/science.1736359]

Table 1. Demographic characteristics and baseline test scores for the participants (n=59)

Characteristics		Mean	Standard deviation
Age		21.41	2.48
Characteristics		**N**	**%**
Gender	**Female**	37	62.7
	Male	22	37.3
Superiority of hand	**Right**	54	91.5
	Left	5	8.5
Characteristics		**Mean**	**Standard deviation**
MoCA score		28.94	0.79
Forward digit span test score		7.22	0.83
Backward digit span test score		5.42	0.69
Word recognition score		97.88	2.81
Word recall score		87.88	5.66
Word judgment score		96.69	4.00
Total WARRM score		94.15	3.47

MoCA; montreal cognitive assessment, WARRM; word auditory recognition and recall measure

Table 2. The mean retest scores for the Persian word auditory recognition and recall measure

Score	Number	Mean	Standard deviation
Word recognition	30	97.50	2.86
Word recall	30	89.83	6.08
Word judgment	30	97.00	4.47
Total	30	94.77	3.75