Auditory and Vestibular Research

The Persian version of the 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.

Romina shahrokhi
https://orcid.org/0009-0005-7478-6573

romina.shahrokhi75@gmail.com
Maryam ramezani
ORCID ID: https://orcid.org/0000-0003-1556-6130
maryamramezani16@gmail.com
Mohanna Javanbakht
ORCID ID: https://orcid.org/0000-0002-2876-3208
mo.javanbakht@uswr.ac.ir
Behrooz Mahmoodi-Bakhtiari
ORCID ID: https://orcid.org/0000-0002-5671-4040
mbakhtiari@ut.ac.ir
Enayatollah Bakhshi
ORCID ID: https://orcid.org/0000-0001-8049-0190
bakhshi@razi.tums.ac.ir

Highlights

- The Persian WARRM test was designed to measure auditory working memory
- The WARRM test showed high validity and reliability based on CVR, CVI, and ICC scores
- WARRM scores showed significant correlations with forward /backward digit span tests

Abstract:

Background and aim: Speech perception in noise involves more than recognizing speech sounds. Given the importance of working memory in speech comprehension and the lack of a Persian working memory test with appropriate lexical load, this study aimed to develop and validate the Persian version of the Word Auditory Recognition and Recall Measure (WARRM) test in 18-25-year-old Persian speakers.

Method: The study involved two main stages: developing the Persian version of the test and evaluating its validity and reliability. Persian words were selected by help of a linguistics expert based on phonetic and semantic features. Face and content validity were confirmed by 10 specialists. Monolingual Persian-speaking students aged 18-25 at Tehran universities in 2023 participated. The WARRM test was administered along with the forward and backward digit span test. Half of the participants repeated the WARRM test after 2 weeks. Data were analyzed.

Result: The WARRM test showed good face and content validity. Overall reliability was 0.72, with subscale reliabilities of 0.49 (recognition), 0.73 (judgment), and 0.87 (recall). The ICC for test-retest was 0.88 (overall), 0.89 (recall subscale), and 0.48 (recognition subscale). Significant positive correlations were found between the WARRM score and forward digit span ($r=0.65$, $p<0.001$) and backward digit span ($r=0.43$, $p<0.001$).

Conclusion: The results indicate that the Persian version of the WARRM has good validity_ reliability, making it suitable for use in clinical and rehabilitation procedure.
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] . Speech perception in noise depends on the interaction of sensory and cognitive processes, involving both bottom-up and top-down factors [۲]. Cognition refers to a wide range of activities performed by the human brain, including perception, thinking, reasoning, remembering, analyzing, planning, attention, mood , combination of ideas, innovation, judgment, awareness, and insight [۳]. Although various cognitive abilities are involved in speech comprehension, much research in this area has focused on the effects of working memory [۴]. Working memory is the ability to actively maintain and manipulate information in the mind for performing complex tasks such as reasoning, comprehension, and learning. In working memory, incoming data is evaluated and manipulated[۵] . Working memory is crucial in speech comprehension, especially in complex auditory situations. The need to assess auditory working memory is felt in the clinical evaluation of hearing[۶] . Its evaluation can be a useful tool in diagnostic and auditory rehabilitation domains.

Various tests exist to assess working memory with different test materials (words, sentences, numbers), different stimuli (visual, auditory), and different response methods (speaking, ordering, pointing). Notable working memory capacity tests validated in Iran include: N-back test[۷, ۸] , reading span tests[9], forward and backward digit span tests[10], and non-word repetition tests[10]. However, some studies have shown that these tests do not correlate well with the ability to understand speech in the presence of noise[11, 12]. Moreover, the observed lack of correlation between speech comprehension and working memory in some studies may be due to inappropriate test materials[12, 7]. Although in 2011, Aghamolaei et al. developed and evaluated a Persian version of a dichotic test Dichotic auditory verbal memory test (DAVMT) to assess auditory-verbal memory, this test is conducted dichotically and does not exclusively assess memory, potentially involving dichotic disorders and corpus callosum issues in the overall test result [13].

Word Auditory Recognition and Recall Measure (WARRM) is a relatively new test designed to simultaneously assess word recognition and auditory working memory performance using word stimuli in adults [14]. Smith et al. also developed an abbreviated version of this test [15]. Their study indicated that both the original and abbreviated versions of WARRM had recall scores, and recognition scores correlated with other memory tests. An auditory working memory test like WARRM might provide a more practical and relevant measure compared with other working memory tests. First, the test duration is short. Second, using speech stimuli for evaluating the interaction between auditory and cognitive processing experienced in everyday listening is more valid [16]. Given the importance of working memory in speech comprehension and past study results indicating that lexical load in test materials can affect outcomes, as confirmed by Kowialiewski et al. [17], and the mentioned benefits of the WARRM test and the lack of a psychometrically validated Persian version, this study aims to Develop and assess the validity and reliability of the Persian version of this test.

Methods

The study consists of two main parts: first, preparation and development of the Persian version of the WARRM test, and second part, execution of the test in 59 normal hearing, Persian speaking participants and comparison of its results with forward and backward digit span tests. The test- retest reliability assessment after two weeks were considered.

Initially, after obtaining permission from the original author, the English version was analyzed by a linguistics expert for phonetic, semantic, and syllabic features. Many monosyllabic words from everyday language and sources like the Academy of Persian Language were gathered and categorized into consonant-vowel, consonant-vowel-consonant, and consonant-vowel-consonant-consonant groups. The excluded criteria were words with multiple meanings, proper nouns (names of places or people), prepositions, verbs, borrowed words from other languages, uncommon or limited usage words, and words with different pronunciations or unstable meanings. A list of 100 balanced monosyllabic words were selected. 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 were balanced based on the phonetic approach with same focus on high-frequency and low-frequency sounds in selected vowels and consonants. For content and face validity evaluation,

an email was sent to 8 audiologist and 2 linguistics with explanations regarding the test and vocabulary selection processing. The face validity of material was also confirmed not only by that 10 specialists but also with participate of 5 person with middle-high school education. For face validity, reviewers rated each word on a scale from "very weak" (1 point) to "very strong" (5 points). Based on Lawshe's method for determining the Content Validity Ratio (CVR), each item in the questionnaire had three options: "essential," "useful but not essential," and "not essential." The experts also assessed the Content Validity Index (CVI) for the characteristic of "relevance" according to the Waltz and Bausell method [22]. For each word, the experts chose one of four options: "not relevant," "somewhat relevant," "relevant," and "highly relevant." The CVR and CVI values for each word were then calculated based on the specialists' opinions. An acceptable CVR value, given the number of participating audiologists in the survey for evaluating the questionnaire's validity, is a value above 0.62, and the acceptable CVI for each question is more than 0.79. After CVR and CVI evaluation for each word, necessary changes were made based on specialist suggestions and sample cases comments. Final selected Words that met all considered criteria were recorded by a female speaker with a clear and familiar voice at Golbang Institute. Words with uncommon pronunciation were replaced. By Audacity software (a sound editing software), the recorded words were categorized into five 20-words lists. Like the English version of the WARRM, each list had one sets of 2, 3, 4, 5, and 6 words (20 words in total). In the execution phase, 59 monolingual Persian-speaking students aged 18-25 from Tehran universities were selected based on entry criteria. The study was conducted as a pilot. Participants filled out consent forms, medical history, and education level questionnaires. After otoscopy, tympanometry, and audiometry confirming normal hearing, participants had to score above 26 in the Montreal Cognitive Assessment (MoCa). Exclusion criteria included unwillingness to continue, poor attention, lack of cooperation, and loss of entry criteria. Additionally, forward and backward digit span tests were calculated to examine their correlation with other working memory tests. The WARRM test was repeated after 2 weeks in half of the initial participants. The data from these two tests were analyzed.

The abbreviated WARRM test was administered using ASUS laptop and Tsco headphones that calibrated with SLM L 2250 which is manufactured by the Danish company (Brüel & Kjær). Each list contained 20 words arranged in sets of 2, 3, 4, 5, and 6 words. Participants first heard a carrier phrase "you will cite" followed by a target word after 50 milliseconds. Participants had 3 seconds to repeat the word before the onset of the next carrier phrase, and 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. A judgment score is calculated in terms of overall percent correct but serves only to ensure that participants were engaged in the task. Thus, judgment scores should be sufficiently high to have confidence that the listener was engaged in the alphabet task and not just guessing. But this judgment task has no effect or representation on the overall WAARM score. Recognition and judgments had to be done quickly as there was a 3-second interval between 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. WARRM provided two scores: word recognition and recall. The recognition score was the percentage of correctly recognized words, and the recall score was the percentage of correctly recalled words which are registered by the examiner. The judgment score regarding whether a letter falls in the first or second half of the alphabet is calculated as a percentage of overall correctness. The WARRM materials were presented using a playback device and delivered binaurally through headphones at 60 dB Finally the reliability of test was examined after two weeks.

Data analysis was conducted using SPSS software version .۱۷ Lawshe's method was used to assess content validity. The Intraclass Correlation Coefficient (ICC) was used to measure test-retest reliability. Pearson's correlation coefficient was used to determine the correlation with other versions of the test. Independent t-tests were used to assess differentiation. Cronbach's alpha was used to evaluate internal consistency reliability.

Results:

Demographic and Basic Information of Study Participant stated in (table 1).Data analysis was based on the fact that the highest possible score on face validity scale is 5 and the average scores given by the judges for all words are above 4 with mean score of 4.۴8±0.۱3, it can be concluded that the Reviewed words has acceptable face validity. The CVR for all words was reported to be above 0.62 with mean score of ۰,۶۸±0.2۷. CVI was reported to be above 0.79 with mean score of ۰,۸۱±0.3۱. Therefore, these words have a good content validity.

To assess the reliability of the WARRM test, the Kuder-Richardson formula was used since the responses were dichotomous. The overall reliability was 0.72. For assessing the reliability of each subscale in relation to the total,

the intraclass reliability coefficient was used, which was calculated to be 0.49 for the recognition subscale, 0.73 for the judgment subscale, and 0.87 for the recall subscale. Additionally, the overall reliability for all three items together, or the entire test, was 0.83. For the test-retest reliability, ICC was used, yielding 0.88 for judgment, 0.89 for recall, and 0.48 for recognition. The ICC for the total test score was 0.92.
In this study, the Spearman correlation coefficient was used. A significant correlation was found between the WARRM scores and both the forward digit span ($r = 0.65$, $p < 0.001$) and the backward digit span ($r = 0.43$, $p = 0.001$). This indicates a direct and significant correlation between the WARRM test scores and both the forward and backward digit spans (Table 2).

Discussion

The main purpose of this study was to develop a set of Persian monosyllabic words for measuring the Auditory working memory in people who speak Persian. To achieve this, five phonetically balanced and equivalent lists of 20 monosyllabic words were developed. Its validity and reliability, including content validity, face validity, test-retest reliability, and internal consistency, were evaluated and achieved satisfactory scores. Data analysis in results part, indicated that the Persian version of the word recognition test had high validity and reliability based on CVR/CVI and ICC/test-retest findings.
The term working memory refers to a brain system that provides temporary storage and manipulation of the information necessary for such complex cognitive tasks as language comprehension, learning, and reasoning [23]. The WARRM test (Word Auditory Recognition and Recall Measure) is a test for assessing working memory capacity, with its English version created in 2016 by Smith et al. [14]. An abbreviated version was developed in 2020, showing high compatibility with the original version [15]. The word lists in the original WARRM test could cause fatigue and cooperation issues among participants, potentially leading to undesirable results for researchers. Therefore, there was a possibility that the use of an extensive vocabulary for assessing speech comprehension and rehabilitation could pose challenges and using abbreviated word lists could play a crucial role in assessing speech comprehension in noise and evaluating participants' working memory.
The study population in the execution part consisted of students, with an average (standard deviation) age of 21.41 (2.48) years. Among the participants, 62.7% were female, and 91.5% were right-handed. These demographic findings align with a typical student population, as the majority of university students are female, and the prevalence of right-handedness is much higher than left-handedness in the general population.
WARRM scores are stated in (table 2) which is similar to smith's et al article, they obtain mean score of recognition 99 and 86 for recall [14].
Three groups of listeners participated in the Smith's et al, 2016 study. The groups included 48 younger listeners with normal hearing defined as pure-tone thresholds ≤20 dB HL for 250 to 8000 Hz (M = 22.8 years, SD = 2.7 years; 17 males, 31 females), 48 older listeners with normal hearing (ONH) defined as pure-tone thresholds ≤25 dB HL at 250 to 3000 Hz (M = 66.9 years, SD = 5.1 years, range = 60 to 84 years; 8 males, 40 females), and 48 older listeners with sensorineural hearing loss. Study reported a significant correlation between WARRM and Digit span sequencing in participants ($r=0.45$) and ($P<0.001$) [14]. Smith, Pichora-fuller, examined correlation between WARRM and reading span. Participants were a group of younger listeners with normal hearing and group of older listeners with hearing loss ($n = 24$). There was a significant correlation between both working memory measures for the participants ($r=0.55$) and ($P<0.001$) [16]. In the differential validity assessment of our study, there was a direct and significant correlation between WARRM scores and both forward and backward digit span ($r = 0.65$ and $r = 0.43$, respectively, $p < 0.001$), which is consistent with mentioned studies.
Based on the above findings, it appears that the Persian version of the WARRM test is likely to yield results similar to previous studies in the field of auditory working memory assessment for the Persian language.
Attention to the importance of cognitive abilities including working memory spans in speech perception in noise is recommended that, the correlation between this valid Persian version of the WARRM test with speech-in-noise comprehension tests be evaluated. Additionally, the use of this test for rehabilitation and monitoring purposes can be examined.
This study had some limitations given that the original version of this test was recently prepared, validated versions in other languages are not available, so we had limitation to compare results with other languages versions. It seems that in the WARRM test, multiple auditory and cognitive structures are involved, and up to now we have limitations in understanding these structures and their exact roles.

Conclusion

The results showed that the Persian version of this test is a valid and reliable tool that can be used as a non-invasive clinical and research test for assessing working memory. The test showed a direct and significant correlation with both forward and backward digit spans, indicating its effective association with auditory working memory. Thus, it can be used alongside existing memory tests as a suitable lexical burden test.

Ethical Considerations:

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 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.

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 Jan 31;255(5044):556-9. [DOI:10.1126/science.1736359]

Table1-Demographic and basic information of study participant			
Age		Mean	standard deviation
		۲۱,۴۱	2.48
Gender	Female	Number	Percentage
		۳۷	62.7%
	Male	۲۲	37.3%
Superiority of hand	Right	۵۴	91.5%
	Left	۵	8.5%
Variable		Mean	standard deviation
Moca score		۲۸,۹۴	۰,۷۹
Forward digit span		۷,۲۲	۰,۸۳
Backward digit span		۵,۴۲	۰,۶۹

Table 2-Word auditory recognition and recall measurement scores and subscales			
WARRM scores	Number	Mean	Standard deviation
Word recognition score	59	۹۷,۸۸	۲,۸۱
Word recall score	59	۸۷,۸۸	۵,۶۶
Word judgment score	59	۹۶,۶۹	۴,۰۰
Total score	59	۹۴,۱۵	۳,۴۷
Word test-retest recognition score	30	۹۷,۵۰	۲,۸۶
Word test-retest recall score	30	۸۹,۸۳	۶,۰۸
Word test-retest judgment score	30	۹۷,۰۰	۴,۴۷
Total test-retest score	30	۹۴,۷۷	۳,۷۵