Original Article

CENTRAL ASIAN JOURNAL of MEDICAL SCIENCES CAJMS

Cent Asian J Med Sci. 2025 Sept;11(3):107-118 https://doi.org/10.24079/CAJMS.2025.03.001

Development and Psychometric Evaluation of Mongolian Monosyllabic Words for Word Recognition Testing

Baigalmaa Oyunbat[1], Urtnasan Davaajav[2], Oyunchimeg Dashnyam[3], Junghwa Bahng[4], Batzorig Bayartsogt[5], Byambasuren Luvsandagva[1]

[1]Department of Otorhinolaryngology, First Central Hospital of Mongolia, Ulaanbaatar, Mongolia;
[2]Institute of Linguistics, Mongolian Academy of Sciences, Ulaanbaatar, Mongolia;
[3]Department of Mathematics, Mongolian National University of Education, Ulaanbaatar, Mongolia;
[4]Department of Audiology and Speech-Language Pathology, Hallym University of Graduate Studies, South Korea;
[5]School of Public Health, Mongolian National University of Medical Sciences, Ulaanbaatar, Mongolia.

Submitted date: June 8, 2025
Accepted date: Sept 22, 2025

Corresponding Author:
Byambasuren Luvsandagva (MD., Ph.D)
Department of Otorhinolaryngology, First Central Hospital of Mongolia, Ulaanbaatar, Mongolia
E-mail: l.byambasuren@fchm.edu.mn
ORCID: https://orcid.org/0000-0003-2814-0534

Objective: To develop a set of Mongolian monosyllabic word lists suitable for measuring Word Recognition Scores (WRS) and to evaluate their validity using psychometric function analysis. **Methods:** In the first phase, 220 Mongolian monosyllabic words that are widely recognized and commonly used in daily conversations by the general population were selected from various linguistic sources and phonemically balanced. A professional female voice artist digitally recorded these words in a sound studio, and their intensity levels were normalized using the Root Mean Square (RMS) value at 1000 Hz. In the second phase, word recognition tests were administered to 50 participants with normal hearing, aged 18 to 29 years. The tests took place in a sound-treated booth, with words presented at intensity levels ranging from –4 to 18 dB HL in 2 dB steps. The 220 words were randomly divided into four blocks of 55 words each. In the third phase, psychometric functions were generated using logistic regression analysis. To determine statistical differences between the lists and half-lists, chi-square (χ^2) tests and one-way analysis of variance (ANOVA) were used. **Results:** From the original pool of 220 words, the 200 most perceptible monosyllabic words were selected based on recognition accuracy across presentation levels. These were divided into four balanced 50-word lists and eight 25-word half-lists. Psychometric functions were calculated for each list using logistic regression. The average slope at the 50% threshold was 10.43%/dB, and the slopes between 20% and 80% recognition ranged from 8.61 to 9.38%/dB. Statistical analyses showed no significant differences in recognition scores or intensity levels across lists and half-lists, confirming their equivalence. Intensity adjustments (≤ ±0.3 dB) were applied to standardize the 50% threshold across all lists (mean=6.16 dB HL). The results demonstrate that the developed Mongolian monosyllabic lists are phonetically and perceptually homogeneous, making them suitable for clinical and research use. **Conclusions:** This study developed a standardized set of Mongolian monosyllabic word lists with consistent psychometric properties. These materials are suitable for use by otologists and audiologists in clinical assessments and speech audiometry research in Mongolia.

Keywords: Speech audiometry, Monosyllabic words, Word recognition, Psychometric function, Intensity level

Introduction

According to the World Health Organization (2021), approximately 430 million people worldwide (around 5% of the global population) require rehabilitation services due to moderate or higher degrees of hearing loss. Projections estimate that by 2050, this number will reach 700 million, meaning that nearly one in ten people globally will have hearing loss.[1,2] Adult-onset hearing loss is considered the second leading cause of Years Lived with a Disability (YLD) and is the fifteenth leading contributor to the Global Burden of Disease (GBD).[3] The impact of hearing loss on an individual largely depends on the extent to which their ability to comprehend speech is affected.[4] Improvement in speech perception is considered one of the primary objectives in the treatment of hearing impairment. Speech audiometry plays a critical role in selecting, planning, implementing, and evaluating appropriate rehabilitation methods.[5]

The initial step in the effective rehabilitation of individuals with hearing impairment is a comprehensive audiological assessment to determine the presence, type, and degree of hearing loss.[2] Although pure-tone testing is a quick and reliable method for measuring frequency-specific aspects of hearing impairment, audiological evaluations are generally considered incomplete without assessing an individual's ability to perceive and process speech. Speech audiometry is a fundamental component of modern hearing diagnostics and clinical audiological research. It serves as a primary tool for determining the degree and type of hearing loss, particularly in the identification of retrocochlear pathologies and auditory processing disorders.[6–8] To ensure accurate assessment in speech audiometry, it is essential to use standardized word materials specifically developed in the listener's native language, with careful consideration of its phonetic and linguistic features.[5,9,10] This assessment serves not only to determine the quality of hearing function but also to evaluate the individual's ability to accurately perceive, discriminate, and comprehend auditory stimuli such as monosyllabic, disyllabic, trisyllabic words, and sentences. As such, it constitutes an integral component of modern auditory diagnostics and clinical audiological research.[11–15]

Speech audiometry uses various types of word materials, and the selection of these materials depends on the type and purpose of the assessment.[8] The most commonly used speech audiometry method is the Word Recognition Score (WRS). By presenting words at intensity levels above the speech recognition threshold, this test provides important information about the listener's ability to understand speech in everyday communicative situations, thereby supporting the evaluation of real-world hearing performance.[12–14] Word recognition tests are generally performed using meaningful words, and word recognition scores are usually expressed as the percentage of correctly identified test words. The psychometric function in word recognition testing illustrates how word recognition scores vary as a function of the presentation level of the test words. Both the recognition score and the psychometric function serve as valuable indices for audiologists in evaluating hearing devices and planning aural rehabilitation strategies.[16–18] The use of appropriate test materials is crucial to ensuring the reliability and diagnostic sensitivity of word recognition tests. Specified criteria for selecting word lists used in Word recognition tests include monosyllabic structure, equivalent average difficulty both between and within lists, equivalent phonetic composition across lists, a composition representative of the spoken language, inclusion of commonly used words, and overall psychometric balance.[19–22] A phonetically balanced list is one in which the occurrence or frequency of the phonemes approximate the language from which the phonemes are derived; in many occasions the syllables contain a consonant-vowel-consonant formation.[23] Another important principle in designing speech test materials is the number of test items.[24,25] Typically, 50-item word lists are adopted in word recognition tests. The inclusion of 50 item increases the reliability of word recognition scores. Nevertheless, audiologists commonly use 25-items word lists to decrease test time and thereby reduce the fatigue effect for participants.[19,25–27]

The first development of speech audiometry materials that reflected the phonetic characteristics of the Mongolian language was conducted in 1988 by Shagdar, et al. and colleagues. Using sonographic analysis, they compiled a mixed list of 10-word groups comprising 20 phonetically balanced words (including monosyllabic, disyllabic, trisyllabic, and numeral words), which were recorded on magnetic tape. In their study, the authors selected thematic texts written by Mongolian authors to represent the richness of the Mongolian vocabulary and analyzed the phoneme and letter frequency in these texts. Based on the results, the final word list included the most frequently occurring phonemes in Mongolian, such as а (11%), э (8%), р (7%), н (6%), л (5%), х (5%), д (5%), г (5%), и (4%), о (4%), and б (4%).[28] Unlike languages with a rich inventory of monosyllabic words, the Mongolian language is characterized by specific phonological features such as vowel harmony, position-sensitive phonotactics, and an agglutinative structure. These linguistic constraints naturally limit both the number and structural variety of monosyllabic words in Mongolian. Common monosyllabic word forms in Mongolian include CV (consonant-vowel), VC (vowel-consonant), and CVC (consonant-vowel-consonant) structures. However, frequency-sensitive phonemes such as fricatives (/x/, /s/, /ʃ/) and nasals (/m/, /n/, /ŋ/)—which are important for detecting hearing loss—occur less frequently in natural monosyllabic vocabulary. Therefore, creating a phonetically balanced and clinically appropriate set of monosyllabic words requires careful selection.[29–31] In this study, we selected and refined a set of 220 monosyllabic words based on these linguistic criteria. The final list serves as a foundational material for a reliable, Mongolian-based WRS test, ensuring both phonetic representativeness and clinical applicability. Accordingly, the purpose of the present study was to develop a set of Mongolian monosyllabic word materials suitable for measuring WRS and to evaluate them by determining their psychometric function curves.

Material and Methods

Research Phases

The following sections describe the three phases of the research methodology: (1) preparation of the test material, (2) selection of homogeneous material for use in the word lists, and (3) construction of the final word lists. Each phase includes a description of the methods and corresponding results.

Phase 1: Development of phonemically balanced word lists in Mongolia

Due to the phonological characteristics of the Mongolian language, the number of naturally occurring monosyllabic words is relatively limited. Therefore, words with simple phonetic structures and frequent usage in everyday language were prioritized. A total of 220 monosyllabic words intended for use in WRS testing were compiled from sources such as school textbooks, dictionaries, and an online Mongolian word corpus.[32,33] General terminology, as well as words containing culturally or religiously sensitive meanings, were excluded.

Word familiarity evaluation

The collected monosyllabic words were rated by 20 participants with varying social and educational backgrounds using a 5-point familiarity scale: 1 – not familiar at all, 2 – slightly familiar, 3 – somewhat familiar, 4 – relatively familiar, and 5 – very familiar (commonly used). Words with an average rating of ≥ 4 were selected, resulting in 220 monosyllabic words being retained for recording and further evaluation.

Content validation of the word list

Based on criteria developed by Shi, et al.[34] five domain experts independently evaluated each word to ensure that it did not contain emotionally, culturally, or religiously sensitive content. As a result, one word was excluded, and 219 words were finalized. Subsequently, phonemic balancing of the selected words was conducted based on phoneme and word frequency data in the Mongolian language to ensure even distribution.[28,35-37]

Digital recording of words

Although male voices have traditionally been preferred in earlier studies to reduce inter-talker variability, recent research demonstrated that speech materials recorded by female talkers also yield clinically acceptable and psychometrically reliable results.[38] Accordingly, the 220 monosyllabic words were recorded in a double-walled recording booth at a professional studio in Ulaanbaatar. The words were articulated by a professional female voice artist who was instructed to produce them with normal vocal effort at a 0° azimuth (directly in front of the microphone) at a distance of 6 cm. This corresponded to an average Sound Pressure Level (SPL) of approximately 65 decibels (dB SPL), as monitored by a sound level meter. Recordings were made at a sampling rate of 44.1 kHz with 24-bit resolution. Each word was recorded three times, and the most clearly articulated version was selected for further use. The Root Mean Square (RMS) amplitude of all recorded words was equalized using the long-term average at 1000 Hz to ensure consistent intensity levels across the entire word list.[39]

Phase 2: Standardization of word lists

Participants

A total of 50 normal-hearing individuals (25 females and 25 males), aged between 18 and 29 years (M=23.5, SD=3.61), participated in the evaluation of the psychometric functions of the test materials. Prior to testing, all participants underwent pure-tone audiometric screening to confirm that their hearing thresholds in both ears were below 20 dB HL (decibel hearing level). Tympanometry was also testing was also conducted to ensure normal middle ear function.[40] The test ear was selected based on the better pure-tone average (PTA) across 500, 1000, and 2000 Hz frequencies. Summary statistics of the participants' hearing thresholds are presented in Table 1.

Table 1. Pure Tone Threshold (dB HL) of 50 Normally Hearing Subjects

Frequency (Hz)	M	Minimum	Maximum	SD
250	2	-5	15	4.5
500	2.6	-5	10	3.5
1000	3.6	-5	15	4.2
2000	0.5	-10	10	5.4
4000	1.4	-10	15	6.1
8000	5.7	-5	20	7.1
PTA[a]	2.23	-6.8	11.7	4.4

[a]PTA - arithmetic average of thresholds at 500, 1000, & 2000 Hz

Evaluation of monosyllabic words

Participants were not familiarized with the monosyllabic words prior to testing. Each participant was presented with 220 monosyllabic words via DD45 headphones in a quiet, sound-treated environment. To generate the psychometric function (PI–PB) across hearing levels, words were presented at intensity levels ranging from –4 dB HL to 18 dB HL in 2 dB increments. The 220 stimuli were randomly divided into four blocks of 55 words each. Both the order of the blocks and the order of words within each block were randomized for each participant, with a 4-second silent interval between words. Participants were encouraged to repeat what they heard, to guess if uncertain, or to say "I don't know" if the word was not heard. Breaks were permitted between blocks if participants experienced fatigue. All responses were automatically recorded on a personal computer for subsequent analysis. Each response was scored as 1 for correct and 0 for incorrect or missing responses. The accuracy of responses was independently verified by three audiologists. A logistic regression analysis was applied to derive the psychometric function for each word.

Statistical Analysis

To analyze the study data, logistic regression, one-way analysis of variance (ANOVA), and chi-square (χ^2) tests were performed. The dependent variable was defined in binary format based on whether the participant correctly identified the word (1 = correct, 0 = incorrect or no response), while the independent variables were the presentation level of the stimulus word and the word list to which it belonged. Based on the response scores for each word in the monosyllabic word lists, logistic regression was applied to estimate the threshold and slope parameters of the psychometric functions. To determine whether statistically significant differences existed between the lists and half-lists, chi-square (χ^2) tests and one-way ANOVA were conducted.

Ethical Statement

This study was approved by the Research Ethics Committee of the Mongolian National University of Medical Sciences (Approval No. 2024/3-06, June 14, 2024). All procedures were conducted in accordance with institutional and national ethical guidelines and adhered to the principles of the Declaration of Helsinki. Participation was voluntary, and written informed consent was obtained from all participants after they were informed about the study's purpose, procedures, and potential benefits. Confidentiality and anonymity were strictly ensured, and no identifying information was included in the analyses or publications.

Results

Standardization of the Developed Word Lists and Assessment Consistency

After compiling the raw data, each monosyllabic word was ranked by difficulty based on how many times it was correctly identified across all intensity levels and test subjects. Words identified more frequently received higher rankings. To remove unfamiliar or less clearly articulated words, the 200 most perceptible words were selected for inclusion in the lists and half-lists. These 200 words were divided into four balanced lists of 50 words each. The lists were created using random block assignment to ensure comparable difficulty levels across them. Specifically, the first four words from the rank-ordered list were randomly assigned, one to each list, and this process was repeated until each list contained 50 words. The four equivalent monosyllabic word lists, recorded by a single talker, are shown in Table 2. After creating the four balanced 50-word lists, eight half-lists of 25 words each were constructed. Each list was split into two half-lists of similar difficulty by grouping the 50 words into 25 consecutive pairs. The first word of each pair was assigned to half-list A, and the second to half-list B. The Mongolian monosyllabic half-lists are presented in Table 3. Following the creation of the monosyllabic lists and half-lists, logistic regression was used to calculate regression slopes and intercepts for each of the four lists and eight half-lists, corresponding to recordings by a single talker. The logistic regression slope and intercept values for each list and half-list are shown in Table 4.

Table 2. Mongolian Monosyllabic Lists in Rank Order from Steepest to Shallowest Slope

List 1		List 2		List 3		List 4	
Багш	Цэр	Эгч	Амт	Шаар	Цонх	Ач	Ноос
Хайч	Од	Ваар	Шал	Өөх	Хана	Гол	Давс
Зах	Гэр	Хайр	Алт	Чих	Найр	Хөрш	Дарс
Жоом	Ой	Тоо	Хийл	Байр	Зар	Навч	Шавьж
Хавч	Хань	Товч	Дүрс	Айл	Зай	Хаш	Угж
Заан	Хөлс	Бийр	Мөр	Морь	Цөс	Тахь	Зун
Хөвч	Үс	Өр	Жил	Эмч	Тос	Шоо	Сүх
Лаа	Дээс	Харц	Бэр	Баг	Хивс	Бар	Хүрз
Бөгж	Сүүл	Цох	Яс	Нуур	Нум	Луу	Лам
Мөөг	Тал	Жор	Нас	Ээж	Тайз	Буга	Дүү
Орц	Яр	Хөл	Унь	Бөх	Үүд	Тор	Цас
Нар	Заль	Оймс	Төл	Арьс	Өвс	Үг	Мөөм
Дуу	Уул	Нүх	Ханш	Хүч	Элс	Цаг	Цус
Талх	Хийд	Гарц	Суга	Зүрх	Жаал	Уур	Дэр
Хүн	Найз	Даага	Хөө	Цоож	Мод	Өмч	Шүүс
Шөл	Гүүр	Толь	Хурд	Хань	Сүнс	Мах	Нүд
Цүнх	Мал	Говь	Зам	Эм	Лааз	Цээж	Саа
Чоно	Зул	Шүүр	Хүж	Муур	Бөөр	Хүү	Цуу
Хаан	Сум	Яам	Хүрд	Данх	Дүн	Өт	Жимс
Хөөс	Бүс	Бал	Дээл	Хот	Бух	Хог	Элч
Иш	Данс	Цамц	Сам	Ам	Уут	Зоорь	Сүү
Боов	Лийр	Ёс	Цув	Хоол	Сэг	Үүл	Сагс
Тоос	Шил	Живх	Сүм	Үнс	Хүйс	Цэг	Цом
Шүд	Нус	Таг	Сойз	Сар	Нэрс	Гинж	Банш
Ингэ	Ус	Цай	Хууль	Дух	Өмд	Мич	Буу

Table 3. Mongolian Monosyllabic Half-lists in Rank Order from Steepest to Shallowest Slope

1A	1B	2A	2B	3A	3B	4A	4B
Яр	Заль	Хүрд	Сойз	Хоол	Дүн	Дарс	Цээж
Бөгж	Ус	Өр	Таг	Өвс	Сар	Элч	Хүрз
Лийр	Мал	Цай	Дээл	Дух	Хана	Гол	Буга
Шил	Ингэ	Жор	Мөр	Баг	Уут	Сүх	Үүл
Жоом	Хийд	Ёс	Живх	Цоож	Ам	Навч	Зоорь
Гүүр	Тоос	Өмд	Бал	Мод	Бөөр	Буу	Шүүс
Талх	Хаан	Амт	Харц	Хот	Нум	Цас	Өт
Од	Чоно	Цув	Шүүр	Тайз	Муур	Хөрш	Нүд
Иш	Цүнх	Товч	Үүд	Үнс	Өөх	Угж	Өмч
Сум	Шөл	Шал	Сүм	Данх	Хонь	Дэр	Мах
Ой	Хүн	Хивс	Яс	Зай	Лааз	Банш	Гинж
Цэр	Тал	Хөл	Гарц	Ээж	Зүрх	Цом	Хумс
Зун	Гэр	Алт	Нүх	Цөс	Жил	Үс	Дүү
Хөлс	Нус	Хөө	Сэг	Сам	Найр	Цэг	Үг
Дуу	Орц	Яам	Ваар	Нуур	Хүйс	Тахь	Бар
Мөөг	Найз	Давс	Хийл	Бөх	Бух	Цаг	Цус
Нар	Зам	Цамц	Цох	Унь	Нэрс	Хүү	Луу
Боов	Лаа	Толь	Бүс	Арьс	Цонх	Шавь	Тор
Шүд	Хөвч	Даага	Сүнс	Морь	Эмч	Шоо	Сагс
Хань	Данс	Элс	Бэр	Нас	Хурд	Сүү	Цуу
Дээс	Суга	Оймс	Ханш	Зар	Айл	Ач	Хаш
Хүж	Уул	Зул	Тоо	Хууль	Байр	Мөөм	Саа
Хөөс	Зах	Сүүл	Хайр	Хүч	Чих	Ноос	Мич
Заан	Хайч	Говь	Төл	Эм	Тос	Уур	Хог
Хавч	Багш	Бийр	Эгч	Жаал	Шаар	Лам	Дүрс

Table 4. Mean Performance of Mongolian Monosyllabic Lists and Half-lists

List / Half-list	a[a]	b[b]	Slope at 50%[c]	Slope at 20-80%[d]	Threshold[e]	ΔdB[f]
1	2.57810	-0.41409	10.4	8.96	6.18	0.02
2	2.50858	-0.42008	10.5	9.10	5.96	-0.20
3	2.74860	-0.43025	10.5	9.30	6.33	0.17
4	2.61353	-0.41292	10.3	8.94	6.18	0.02
M	2.61	-0.42	10.43	9.07	6.16	0.00
Minimum	2.51	-0.43	10.32	8.94	5.96	-0.20
Maximum	2.75	-0.41	10.52	9.30	6.33	0.17
Range	0.24	0.02	0.20	0.37	0.36	0.36
SD	0.09	0.01	0.09	0.15	0.13	0.13
1A	2.6418	-0.4156	10.39	8.99	6.25	0.09
1B	2.514	-0.413	10.31	8.93	6.11	-0.05
2A	2.4447	-0.4158	10.42	9.02	5.88	-0.23
2B	2.572	-0.424	10.61	9.18	6.04	-0.12
3A	2.8216	-0.4345	10.39	9.38	6.48	0.32
3B	2.6756	-0.4260	10.65	9.22	6.17	0.01
4A	2.4525	-0.3977	9.94	8.61	6.07	-0.09
4B	2.7746	-0.4281	10.7	9.26	6.29	0.13
M	2.61	-0.42	10.43	9.07	6.16	0.00
Minimum	2.44	-0.43	9.94	8.61	5.88	-0.28
Maximum	2.82	-0.40	10.70	9.38	6.48	0.32
Range	0.38	0.04	0.76	0.78	0.60	0.60
SD	0.13	0.01	0.23	0.23	0.17	0.17

[a]a - regression intercept. [b]b - regression slope. [c]Psychometric function slope (%/dB) at 50% was calculated from 49.999 to 50.001%. [d]Psychometric function slope (%/dB) from 20–80%. [e]Intensity required for 50% intelligibility. [f]Change in intensity required to adjust the threshold of the list to the mean threshold of four lists and eight half-lists (6.16 dB HL).

In Equation 1, p represents the proportion of correct recognition, a is the regression intercept, b is the regression slope, and i is the presentation intensity level in dB HL. This logistic regression model was used to generate psychometric functions across intensity levels.

$$p = \frac{1}{1 + \exp(-(a + b \cdot i))} \times 100$$

The predicted percentages of correct word recognition were calculated for each monosyllabic word list and half-list across a presentation range from –4 to 18 dB HL in 2 dB increments. Using these values, psychometric functions were constructed for each word list. Subsequently, the threshold (defined as the intensity level required to achieve 50% word recognition), the slope at the threshold, and the slope between 20% and 80% recognition were calculated by substituting the corresponding proportions into the inverse logistic equation (Equation 2).

$$i = \frac{\ln\left(\frac{p}{1-p}\right) - a}{b}$$

In Equation 2, i represents the presentation intensity level in dB HL, p is the proportion of correct recognition, a is the regression intercept, and b is the regression slope. The psychometric function slopes at the 50% location for the monosyllabic lists and half-lists ranged from 9.94 to 10.70%/dB (M=10.43 %/dB) for the talker's recordings. When measured between the 20% and 80% points of the functions, the slopes were slightly lower, ranging from 8.61 to 9.38%/dB (M = 9.07%/dB), compared with the 50% threshold midpoint. Logistic regression analysis was then performed to model the relationship between presentation intensity and word recognition. To examine list equivalence, a two-way chi-square (χ^2) test was conducted with intensity and list as independent variables and response as the dependent variable. No statistically significant differences were found among the four 50-word lists, $\chi^2(3) = 0.12$, $p = .98$. A one-way analysis of variance (ANOVA) comparing mean intensity levels likewise revealed no significant differences, F = 0.004, $p = .9996$, confirming equivalence in both recognition performance and intensity presentation. Similarly, the eight 25-word half-lists were evaluated using the same statistical methods. The two-way chi-square test revealed no significant differences in recognition performance across half-lists, $\chi^2(7) = 2.63$, $p = .918$, and the one-way ANOVA showed no significant differences in mean intensity levels, F = 0.036, $p = .9997$. These findings suggest that both the complete lists and half-lists were psychometrically equivalent in terms of word recognition and presentation intensity.

The intensity of each word in the Mongolian monosyllabic four lists and eight half-lists was digitally adjusted so that the 50% threshold of each list matched the overall mean threshold of the four lists and eight half-lists (6.16 dB HL). The adjustments applied to each word are summarized in Table 4. The psychometric functions of the monosyllabic lists and half-lists before and after adjustment are shown in Figures 1 and 2, respectively. Figure 3 presents the mean psychometric functions for the combined complete lists and half-lists, both before and after normalization. After intensity adjustment, the predicted psychometric functions were highly consistent across all lists, requiring only minor modifications (≤ ±0.3 dB) to equate the 50% recognition thresholds. These results demonstrate that the lists are both phonetically and perceptually balanced, thereby supporting their equivalence and validating their use in clinical speech audiometry.

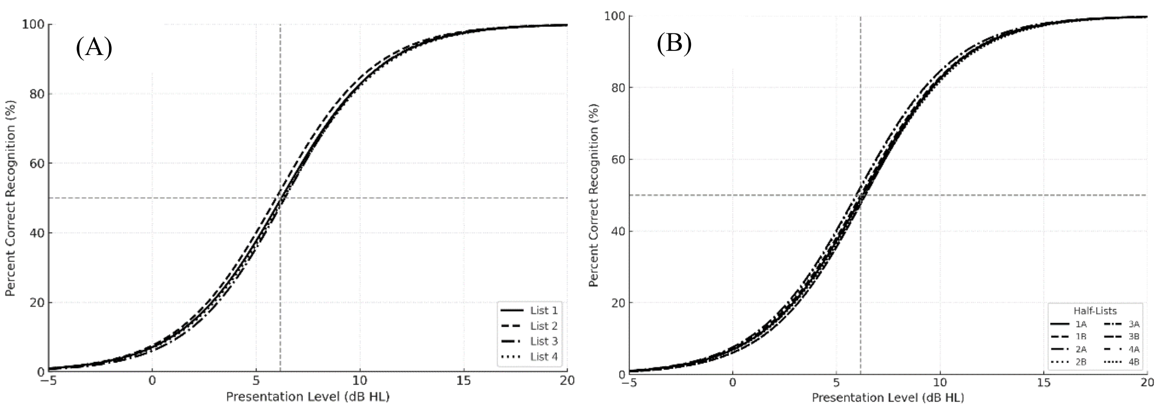

Figure 1. (A) Psychometric functions for the four Mongolian monosyllabic lists recorded by a single talker before intensity adjustments. (B) Psychometric functions for the eight Mongolian monosyllabic half-lists recorded under the same condition.

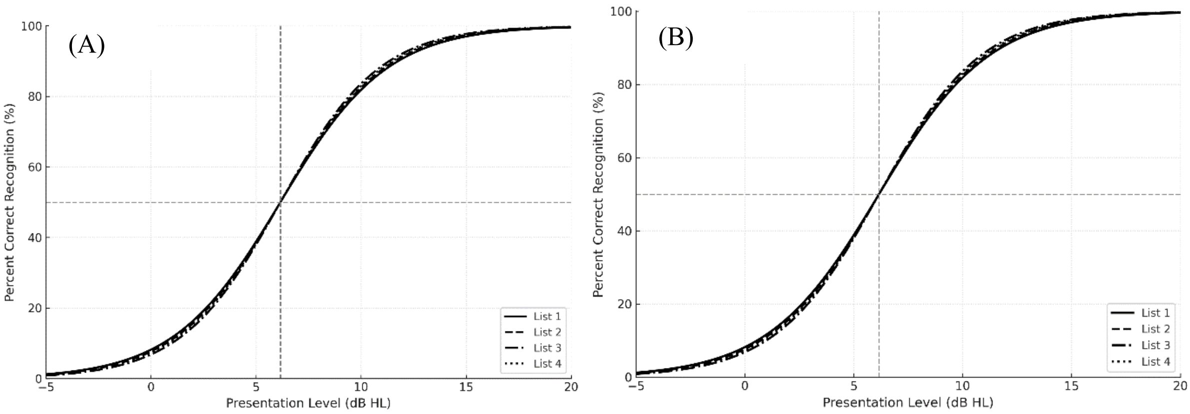

Figure 2. (A) Psychometric functions for the four monosyllabic lists recorded by a single talker, after intensity adjustments to yield 50% recognition performance at 6.16 dB HL. (B) Psychometric functions for the eight monosyllabic half-lists recorded under the same conditions, demonstrating equivalent 50% recognition thresholds at 6.16 dB HL.

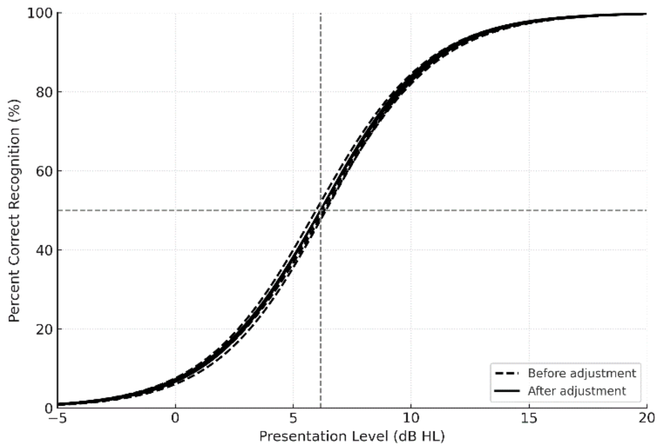

Figure 3. Mean psychometric functions for monosyllabic word lists recorded by a single talker, before and after intensity adjustment. Adjustments were applied to each list and half-list to equalize the 50% recognition threshold at 6.16 dB HL.

Discussion

The main purpose of this study was to develop a set of homogeneous Mongolian monosyllabic word lists for assessing auditory word recognition. Inspection of Figures 1–3 indicates that the developed lists and half-lists demonstrated highly homogeneous performance with respect to audibility and the slopes of their psychometric functions. To examine whether statistically significant differences existed between the monosyllabic lists and half-lists, chi-square tests and one-way analysis of variance (ANOVA) were conducted. As shown in Table 4, the analyses revealed no statistically significant differences among the four full lists or the eight half-lists.

In the present study, all monosyllabic words were recorded by a single female speaker. While this could be seen as a limitation regarding generalizability, previous research has consistently shown that there are no statistically significant differences in recognition performance between male and female speakers.[41-44] Moreover, several studies suggest that female voices often produce more consistent results in clinical and psychometric settings, as their acoustic features (e.g., clearer formant structures and higher fundamental frequency) are more easily distinguishable.[45] Therefore, for this study, using a single female speaker was considered both sufficient and appropriate to ensure list equivalence, especially since psychometric consistency, phonetic balance, and word familiarity were prioritized over speaker variation. As a result, the current study focused on maintaining psychometric equivalence, phonetic balance, and word familiarity across the lists rather than emphasizing talker gender differences. Given the similar recognition performance reported in the literature regardless of speaker gender, using a single female speaker in this study is deemed adequate for ensuring list equivalence.

The psychometric function slopes at 50% for the monosyllabic lists and half-lists ranged from 9. 94 to 10. 70 %/dB (M = 10. 43 %/dB), as shown in Table 4. The slopes within the 20–80% performance range varied from 8. 61 to 9. 38 %/dB (M = 9. 07 %/dB). These results suggest steeper psychometric slopes compared to those reported in most word recognition test materials developed in other languages. For example, English word recognition materials like NU-6 and CID W-22 have shown average slopes of 4. 2%/dB and 4. 6%/dB, respectively; Korean monosyllabic lists displayed slopes of 5. 0%/dB (male) and 5. 1%/dB (female); and Russian lists yielded 5. 8%/dB and 5. 6%/dB. Conversely, a study of Persian monosyllabic word lists reported an average psychometric slope of 20. 20.42%/dB for words recorded by a male speaker, and 19. 24%/dB for those recorded by a female speaker. Additionally, Mahdavi, et al.'s[46] research indicated that thresholds and psychometric slopes for monosyllabic CVC words ranged from 3. 5 to 10. 7 dB HL and from 6. 2%/dB to 4. 1%/dB, respectively, aligning with our findings. Variations in psychometric properties across studies can stem from several factors, including presentation level step size, calibration methods, and syllable structure of the test words.[41,46,47] Compared to other studies, the relatively steep psychometric slopes (M = 9. 85%/dB) noted in the Mongolian monosyllabic word lists may partially reflect the phonological features of Mongolian. Most Mongolian monosyllabic words follow a CVC (Consonant–Vowel–Consonant) structure, with clearly articulated final consonants (such as /n/, /g/, /l/, /m/), which enhance speech clarity and auditory distinction. Moreover, Mongolian root words tend to be morphologically simple and phonetically salient, aiding in higher word recognition rates in suprathreshold conditions. In comparison to languages with more diverse syllable structures, this linguistic uniformity and acoustic clarity likely contribute to steeper psychometric slopes, as participants could detect and recognize words more reliably at lower intensity levels. Thus, the results support the idea that language-specific phonetic and structural features can significantly shape the form and slope of psychometric functions in word recognition testing.

In summary, our study successfully created and validated a set of 200 Mongolian monosyllabic words that showed consistent psychometric properties. These lists were phonetically balanced, perceptually reliable, and statistically comparable between full and half-lists. The materials developed offer standardized tools for clinical audiological assessments and speech perception research in Mongolia. Future research should explore word recognition performance in noise, such as multi-talker babble, at various SNR levels.

Conflict of Interest

The authors state no conflict of interest.

Funding

The Raphael International supported this research through the Korea International Cooperation Agency (KOICA).

Acknowledgements

We would like to express our sincere gratitude to Raphael International and the Department of Speech-Language Pathology at Hallym University, Republic of Korea, for their invaluable collaboration and academic support. We also extend our appreciation to the audiologists and colleagues who assisted in the preliminary word selection and provided professional feedback throughout the study. Finally, we gratefully acknowledge the organizations and professionals who offered constructive comments and advice that significantly improved this research.

Authors Contribution

Baigalmaa Oyunbat: conceptualization, data curation, formal analysis, investigation,
methodology, visualization, writing original draft
Urtnasan Davaajav: methodology, validation
Oyunchimeg Dashnyam: data curation, formal analysis, visualization

Junghwa Bahng: conceptualization, data curation, funding acquisition, methodology, recources,
supervision, validation, writing review & editing
Batzorig Bayartsogt: data curation
Byambasuren Luvsandagva: conceptualization, funding acquisition, project administration

References

1. World Health Organization. *World Report on Hearing*. World Health Organization; 2021. https://www.who.int/publications/i/item/world-report-on-hearing
2. World Health Organization. *Primary Ear and Hearing Care Training Manual*. World Health Organization; 2023. https://www.who.int/publications/i/item/9789240069152
3. Haile LM, Kamenov K, Briant PS, et al. Hearing loss prevalence and years lived with disability, 1990-2019: Findings from the Global Burden of Disease Study 2019. *Lancet*. 2021;397(10278):996-1009. https://doi.org/10.1016/S0140-6736(21)00516-X
4. Abdulhaq NMA. *Speech Perception Test for Jordanian Arabic Speaking Children*. University of Florida; 2006.
5. Noor H, Arif M. Development and validation of phonetically balanced speech perception test in Urdu language. I*nternet J Allied Heal Sci Pract*. 2018;16(4):12-16. https://doi.org/10.46743/1540-580X/2018.1746
6. Naqvi, Yasmin, Sutton AE. *Speech Audiometry*. StatPearls Publishing. Published 2025. https://www.ncbi.nlm.nih.gov/books/NBK587357/
7. Katz J, Chasin M, English K, et al. *Handbook of Clinical Audiology*. 7th edition. Wolters Kluwer Health; 2015.
8. British Society of Audiology. P*ractice Guidance - An Overview of Current Management of Auditory Processing Disorder (APD)*. British Society of Audiology; 2011.
9. Hanson C. Development of Speech Recognition Threshold and Word Recognition Materials for Native Vietnamese Speakers. *BYU Sch Arch*. 2014;4330:1-93. http://scholarsarchive.byu.edu/etd/4330/
10. Harris R W, Nielson W S, McPherson D L. et al. Psychometrically equivalent Polish bisyllabic words spoken by male and female talkers. *Audiofonologia*. 2004;25(1):1-15.
11. Kramer S, Brown DK. Audiology: *Science to Practice, Third Edition*. 3rd ed. Plural Publishing; 2018.
12. Egan JJ. Basic aspects of speech audiometry. *Ear, Nose, Throat J*. 1979;58(5):190–193.
13. Hagerman B. Some aspects of methodology in speech audiometry. *Scand Audiol Suppl*. 1984;(21):1–25. https://pubmed.ncbi.nlm.nih.gov/6589731/
14. Jerger S, Jerger J, Abrams S. Speech audiometry in the young child. *Ear Hear*. 1983;4(2):56–66. https://doi.org/10.1097/00003446-198303000-00003
15. Brad A. Stach VR. *Clinical Audiology: An Introduction, 3rd Edition*. third edit. Plural Publishing; 2022.
16. Dowell RC, Dettman SJ, Hill K, et al. Speech perception outcomes in older children who use multichannel cochlear implants: Older is not always poorer. *Ann Otol Rhinol Laryngol*. 2002;111(5 II):97-101. https://doi.org/10.1177/00034894021110s520
17. Kruger B, Kruger F M. *Speech audiometry in the USA. In: Martin M, ed. Speech Audiometry. 2nd ed*. Whurr; 1997:233–277.
18. Sarant, J Z, Blamey P J, Dowell R C. et al. Variation in speech perception scores among children with cochlear implants. *Ear Hear*. 2001;22(1):18-28. https://doi.org/10.1177/109625060300700101
19. Egan JP. Articulation testing methods. *Laryngoscope.* 1948;58(9):955–991. https://doi.org/10.1288/00005537-194809000-00002
20. Tiravanitchakul R, Kosawat K, Munthuli A. et al. Phonetically balanced and psychometrically equivalent monosyllabic word lists for word recognition testing in Thai. *J Acoust Soc Am*. 2017;138(3):1831. https://doi.org/10.1121/2.0000496
21. Vainutienė V, Ivaška J, Kardelis V. et al. Speech Audiometry: The Development of Lithuanian Bisyllabic Phonemically Balanced Word Lists for Evaluation of Speech Recognition. *Appl Sci*. 2024;14(7):2897. https://doi.org/10.3390/app14072897
22. MacPherson A, Akeroyd MA. Variations in the slope of the psychometric functions for speech intelligibility: A systematic survey. *Trends Hear*. 2014;18:2331216514537722. https://doi.org/10.1177/2331216514537722
23. Brandy WT. *Speech audiometry. Handbook of Clinical Audiology. 5th ed*. Philadelphia: Lippincott Williams & Wilkins. 2002:96-110.
24. Tsai K S, Tseng L H, Wu C J. et al. Development of a Mandarin monosyllable recognition test. *Ear Hear.* 2009;30(1):90-99. https://doi.org/10.1097/AUD.0b013e31818f28a6
25. Han D, Wang S, Zhang H, et al. Development of Mandarin monosyllabic speech test materials in

China. *Int J Audiol*. 2009;48(5):300-311. https://doi.org/10.1080/14992020802607456
26. Lehiste I, Peterson GE. Linguistic Considerations in the Study of Speech Intelligibility. *J Acoust Soc Am*. 1959;31(3):280-286. https://doi.org/10.1121/1.1907713
27. Hirsh I J, Davis H, Silverman S R. et al. Development of materials for speech audiometry. *J Speech Hear Disord*. 1952;17(2):321-337. https://doi.org/10.1044/jshd.1703.321
28. Shagdar L. *Kliniko-Audiologicheskie Aspekty Timpanoplastiki v Usloviyakh Mongol'skoy Narodnoy Respubliki*. Ulaanbaatar / Kiev. 1988.
29. Poppe N. *Mongolian Language Handbook*. Center for Applied Linguistics. 1970.
30. Svantesson J O, Tsendina A, Karlsson A. et al. *The Phonology of Mongolian*. Oxford University Press. 2005.
31. Sanders A J, Bat-Ireedui ZH. *Colloquial Mongolian: The Complete Course for Beginners.* Routledge. 1999.
32. Ashgaa Y. Монгол Хэлний Сурагчийн Тайлбар Толь Бичиг. 2015.
33. Engine S. Mongolian Web Corpus. I*nstitute of the Czech National Corpus*. Published 2016. https://www.sketchengine.eu/mongolian-wac-corpus/
34. Shi J, Mo X, Sun Z. Content validity index in scale development. *J Cent South Univ Medical Sci*. 2012;37(2):152-155. https://doi.org/10.3969/j.issn.1672-7347.2012.02.007
35. Заяа.М. Монгол Үгийн Шуугиант Авианы Энергийн Нягтралыг Судалж, Дижитал Сонсголын Аппарат Тохируулах нь. *Mongolian National University of Medical Sciences*. 2010.
36. Tian J. Exploring the Mongolian spoken language corpus: Mongolian vowels phonetic experimental research. *Commun Humanit Res*. 2023;20:200-204. https://doi.org/10.54254/2753-7064/20/20231349
37. A A. Орчин цагийн монгол хэлний авиа зүй. Scribd. https://www.scribd.com/document/457510298/530382b9de51b20a4841e167f7fae5a5acd35676-pdf
38. Harris R W, Nissen S L, Pola M G.et al. Psychometrically equivalent Russian speech audiometry materials by male and female talkers. *Int J Audiol*. 2007;46(1):47–66. https://doi.org/10.1080/14992020601058117
39. Bilger R C, Matthies M L, Meyer T A. et al. Psychometric Equivalence of Recorded Spondaic Words as Test Items. J Speech, *Lang Hear Res*. 1998;41:516-526. https://doi.org/10.1044/jslhr.4103.516
40. American Speech-Language-Hearing Association. Guidelines for Audiologic Screening. Vol 32;1997. https://doi.org/10.1044/policy.GL1997-00199
41. Martin F N, Champlin C A, Perez D D. The question of phonetic balance in word recognition testing. J Am Acad Audiol. 2000;11(9):489–493. https://doi.org/10.1055/s-0042-1748141
42. Mosleh M. Development and evaluation of a speech recognition test for Persian speaking adults [in Persian]. *Audiology*. 2001;9(1-2):72–76.
43. Harris R W, Kim E, Eggett D L. Psychometrically Equivalent Korean Monosyllabic Speech Discrimination Materials Spoken by Male and Female Talkers. 언어청각장애연구. 2003;8(1):217–243. https://www.e-csd.org/upload/8(1)_12.pdf
44. Chan M, Wong J C, Mok LLN. Development of a Cantonese monosyllabic word recognition test. *Int J Audiol.* 2016;55(9):565–571.
45. Haleh Hassani, Mohsen Ahadi, Farnoush Jarollahi SJ. Development of Persian Monosyllabic and Disyllabic Words for Auditory Test of Adults and Evaluation of Their Face Validity Using Psychometric Function. *Audit vestibulary Res*. 2024;33:202-207. https://doi.org/10.18502/avr.v33i3.15501
46. Mohammad Ebrahim Mahdavi AR. Psychometric function characteristics of Persian consonant-vowel-consonant words. *Audiol Vestib Res*. 2021;30(1):50-55. https://doi.org/10.18502/avr.v30i1.5311
47. Wilson R H, Carter AS. Relation between slopes of word recognition psychometric functions and homogeneity of the stimulus materials. *J Am Acad Audiol. 2001;12(1):7-14.* https://doi.org/10.1055/s-0041-1741115