ORIGINAL ARTICLE
J Audiol Otol 2015;19(2):68-73

pISSN 2384-1621 / eISSN 2384-1710
http://dx.doi.org/10.7874/jao.2015.19.2.68

Test-Retest Reliability of Word Recognition Score Using Korean Standard Monosyllabic Word Lists for Adults as a Function of the Number of Test Words

Jinsook Kim[1], Junghak Lee[2,3], Kyoung Won Lee[2], Junghwa Bahng[2], Jae Hee Lee[2], Chul-Hee Choi[4], Soo Jin Cho[5], Eun Yeong Shin[6], and Jeonghye Park[3]

[1]Division of Speech Pathology and Audiology, Hallym University, Chuncheon,
[2]Department of Audiology, Hallym University of Graduate Studies, Seoul,
[3]Institute of Audiology, Hallym University of Graduate Studies, Seoul,
[4]Department of Audiology and Speech-Language Pathology, Catholic University of Daegu, Gyeongsan,
[5]Department of Speech-Language Pathology and Audiology, Nambu University, Gwangju,
[6]Department of Speech-Language Pathology and Audiology, Sehan University, Mokpo, Korea

Received February 9, 2015
Revised April 11, 2015
Accepted May 14, 2015

Address for correspondence
Junghak Lee, FAAA, CCC-A, PhD
Department of Audiology,
Hallym University of
Graduate Studies,
405 Yeoksam-ro, Gangnam-gu,
Seoul 06198, Korea
Tel +82-2-2051-4950
Fax +82-2-3453-7833
E-mail leejh@hallym.ac.kr

Background and Objectives: The purpose was to establish the test-retest reliability of word recognition score (WRS) using Korean standard monosyllabic word lists for adults (KS-MWL-A) recently developed based on the international standard for speech audiometry (ISO 8253-3:2012). **Subjects and Methods**: Subjects consisted of 159 adults aged to 18 to 25 years with normal hearing sensitivity. WRSs were obtained in 2 dB steps from the level of speech recognition thresholds to the level of 86% correct responses or greater. After one or two weeks, retest was performed. Correlation, confidence interval (CI) and prediction interval (PI) were calculated for the reliability. **Results**: Correlation coefficients were 0.88 for 50 test words, 0.76 for 25 and 0.61 for 10 words. Results also showed that 95% CIs and PIs were narrower for 25 and 50 test words than those for 10 test words. **Conclusions**: Korean WRS using the KS-MWL-A has high reliability for 25 and 50 test words, but relatively low for 10 words. It suggested that 95% CIs for each test words would be criteria for significant differences in WRS for groups and 95% PIs at each score of WRS could be utilized for a considerable difference for each individual at retest. **J Audiol Otol 2015;19(2):68-73**

KEY WORDS: Word recognition score (WRS) · Korean standard monosyllabic word list for adults (KS-MWL-A) · Test-retest reliability · Confidence interval (CI) · Prediction interval (PI).

Introduction

Word recognition score (WRS) is one of the most frequently used measures for speech audiometry. Generally, several monosyllabic word lists (MWL) with a similar level of difficulty are used to get the WRS. Korean MWLs for adults (MWL-A) were recently developed [1] and selected as a Korean standard (KS) for speech audiometry [2]. The KS-MWL-A is widely used in many hearing clinics, hearing aid centers, and auditory rehabilitation centers in Korea. In the clinical settings, WRS gives valuable information to see how much improvement occurred for each individual at the end of treatment, hearing aid fitting, aural rehabilitation, etc. [3-5]. We would not be sure whether the improvement is significant or not, however, if test-retest reliability is not established, which refers to the repeatability of a measure [3-12]. It is well known that parameters affecting WRS include a number of test words, stimulus presentation level and mode, difficulty level of word lists, etc. Although few studies [1,3,12,13] examined test-retest reliability of Korean WRS for adults, their data were not enough to clearly interpret retest results of the KS-MWL-A with respect to aforementioned parameters, be-

cause of differences between old and newly developed word lists, small number of subjects, skewed distribution of WRSs, or homogeneity problem in age.

Indices to show test-retest results include correlation, confidence interval (CI), and prediction interval (PI) in this study. The CI can be described as an estimate of the interval in which the sample mean represents the population mean and the PI as an estimate of the interval in which the retest results will fall with a certain probability, given the results at the previous test [3,8,9]. The PI is useful for making inferences whether the degree of change in WRS at the retest is significant or not for each individual. Therefore, this study tried to investigate the test-retest reliability of the KS-WRS-A according to the recommendations of both international and Korean standards for speech audiometry [2,14]. More specifically, first, correlations between test and retest results were analyzed as a function of the number of test words. Second, CIs were calculated with respect to the whole range of WRS for interpreting group data and finally, PIs were obtained at each score of WRS for clinically interpreting individual retest results.

Subjects and Methods

Subjects

One hundred fifty-nine adults all over the country participated in this study, aged from 18 to 25 years with normal hearing. All subjects were native Korean speakers and had pure tone hearing thresholds equal to or less than 20 dB HL for octave frequencies from 250 to 8000 Hz. Each participant also had A-ype tympanogram and no medical history related with ear. They agreed on and signed in the informed consent form at the beginning of experiment.

Stimulus materials

Four lists of KS-MWL-A were used for measuring WRS which consisted of 200 monosyllabic words (Table 1). Each list has 50 words recorded by a native Korean speaker who was a professional announcer. The monosyllabic words were selected based on word familiarity, phonetical dissimilarity, normal sampling of Korean speech sounds, and homogeneity with respect to intelligibility [1]. Thirty-six bisyllabic words updated by Cho, et al. [15] recorded by a native Korean speaker were used for testing speech recognition threshold (SRT). The recorded speech stimuli were calibrated in reference to a 1000 Hz tone recorded on the compact disc, and the speech stimuli were presented within ±2 dB with respect to the volume unit meter of the audiometer.

Procedure

The GSI 61 audiometer, TDH 50 headphones, and GSI 38 middle ear analyzer were used for this study. Pure tone thresholds were measured from 250 to 8000 Hz in 5 dB steps. According to the pure tone threshold averages (0.5 k, 1 k, 2 k) for each subject, the better ear was selected for measuring SRT and WRS. The SRT was defined as the level necessary for 50% correct responses. Considering the international standard for speech audiometry [14] describing "the test-retest reliability shall be specified for the speech recognition scores 50%, 60%, 70%, 80%, and 90%", WRS was obtained using one of four lists of KS-MWL-A which were randomly presented to each listener beginning at the SRT level. For the above 5 scores, WRS bands consisted of 45−55%, 56−65%, 66−75%, 76−85%, and 86−100% respectively. If WRS at SRT level was equal to or less than 55%, the presentation level ascended from 2 dB above the SRT to the level up to the correct response of 86% or greater in 2 dB steps. If WRS at SRT level was greater than 55%, the presentation level descended to the level below the SRT level in 2 dB steps until the correct response was equal to or less than 55% and then ascended from 2 dB above the SRT level to the level up to 86% or greater in 2 dB steps. Subjects were instructed to repeat each word or to guess if they were unsure. The scoring procedure was to count each of the 50 words as either correctly or incorrectly repeated at each presentation level for each subject. From these data, the percentage of correct responses was computed at each test level as a psychometric function for each subject. After one or two weeks, WRS was retested under the same condition as the first test.

Data analysis

After the raw data were collected, test and retest WRS scores for all subjects were analyzed by Pearson correlation analysis and 95% CIs for 50 words, and first 25 words and first 10 words of each list, respectively, using the statistical package for social sciences (SPSS, version 18, SPSS Inc., Chicago, IL, USA). We also performed one-way analysis of variance and post hoc tests to compare the results of each number of test words.

The CI was obtained from the standard error of mean (SE) which was calculated by dividing the standard deviation (SD) of differences in WRS between test and retest by the root of subject number. The 95% CI was computed by ±2 SE for the whole range of WRS as a function of the number of test items. The PI was determined based on the standard error of measurement (SEM) for each band of WRS which included 45−55%, 56−65%, 66−75%, 76−85%, and 86−100%. To get SEM, SD of differences in WRS between test and retest

Table 1. Korean standard monosyllabic word lists for adults and international phonetic alphabets

List 1			List 2			List 3			List 4		
귀 [kwi]	글 [kɯl]	꽃 [k*ot]	난 [nan]	코 [kho]	양 [jaŋ]	국 [kuk]	십 [sip]	당 [taŋ]	산 [san]	한 [han]	돈 [ton]
남 [nam]	용 [joŋ]	연 [yʌn]	위 [wi]	숯 [sut]	회 [hwɛ]	마 [ma]	얼 [ʌl]	매 [mæ]	두 [tu]	포 [pho]	엿 [jʌt]
해 [hæ]	겉 [kʌt]	달 [tal]	죽 [tɕuk]	오 [o]	겹 [kjʌp]	봄 [pom]	쥐 [tɕy]	신 [sin]	공 [koŋ]	의 [ɯi]	꿀 [k*ul]
밀 [mil]	다 [da]	혀 [hjʌ]	더 [tʌ]	강 [kaŋ]	인 [in]	이 [i]	또 [t*o]	뼈 [p*jʌ]	말 [mal]	섬 [sʌm]	막 [mak]
옷 [ot]	뜰 [t*ɯl]	녹 [nok]	값 [kap]	외 [wɛ]	답 [tap]	농 [noŋ]	파 [pa]	낫 [nat]	넋 [nʌk]	은 [ɯn]	비 [pi]
잔 [tɕan]	피 [phi]	김 [kim]	모 [mo]	장 [tɕaŋ]	노 [no]	학 [hak]	너 [nʌ]	교 [kjo]	방 [paŋ]	때 [t*æ]	궁 [kuŋ]
댁 [tɕek]	상 [saŋ]	약 [jak]	금 [kɯm]	대 [tæ]	불 [pul]	들 [tɯl]	맛 [mat]	발 [pal]	힘 [him]	사 [sa]	단 [tan]
겁 [kʌp]	네 [ne]	덕 [tʌk]	효 [hjo]	육 [juk]	목 [mok]	간 [kan]	끼 [k*i]	저 [tɕʌ]	야 [ja]	집 [tɕip]	요 [jo]
시 [si]	벌 [pʌl]	조 [tɕo]	성 [sʌŋ]	미 [mi]	계 [kjɛ]	컵 [khʌp]	틀 [thɯl]	굴 [kul]	짐 [tɕim]	배 [pæ]	빵 [p*aŋ]
병 [pjʌŋ]	추 [tɕ^{h}u]	군 [kun]	빛 [pit]	솔 [sol]	실 [sil]	새 [sæ]	나 [na]	꾀 [k*ø]	그 [kɯ]	솜 [som]	늪 [nɯp]
소 [so]	만 [man]	잎 [ip]	서 [sʌ]	벼 [pjʌ]	애 [æ]	등 [tɯŋ]	곳 [kot]	살 [sal]	차 [tɕha]	화 [hwa]	옥 [ok]
점 [tɕʌm]	죄 [tɕø]	폐 [phjɛ]	날 [nal]	담 [tam]	널 [nʌl]	개 [kæ]	운 [un]	님 [nim]	읍 [ɯp]	절 [tɕʌl]	재 [tɕæ]
키 [khi]	일 [il]	꿈 [k*um]	깨 [k*æ]	처 [tɕ^{h}ʌ]	안 [an]	징 [tɕiŋ]	덤 [tʌm]	왕 [waŋ]	수 [su]	뇌 [nø]	탈 [thal]
앞 [ap]	구 [ku]	터 [thʌ]	잠 [tɕam]	강 [kaŋ]	쑥 [s*uk]	손 [son]	주 [tɕu]	곰 [kom]	돌 [tol]	귤 [kjul]	기 [ki]
무 [mu]	삼 [sam]	샘 [sɶm]	표 [phjo]	띠 [t*i]	종 [tɕoŋ]	유 [ju]	침 [tɕhim]	면 [mjʌn]	넷 [net]	여 [jʌ]	멋 [mʌt]
논 [non]	도 [to]	능 [nɯŋ]	눈 [nun]	전 [tɕʌn]	음 [ɯm]	밤 [pam]	억 [ʌk]	열 [jʌl]	검 [kʌm]	칼 [kal]	우 [u]
자 [tɕa]	알 [al]		길 [kil]	늘 [nɯl]		예 [jɛ]	후 [hu]		놀 [nol]	씨 [s*i]	

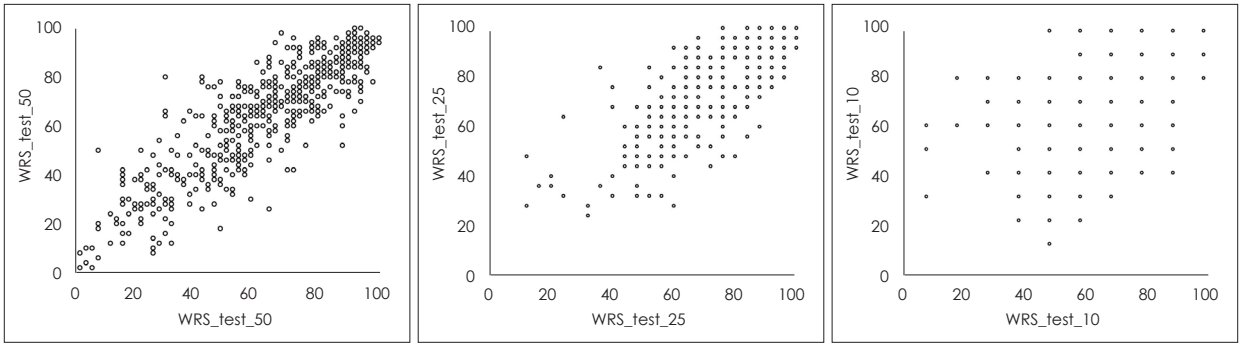

Fig. 1. Scatter plots of WRSs at test and retest for 50 (top), 25 (middle) and 10 (bottom) test words. WRS: word recognition score.

was divided by √2 suggested at previous researches [5,6,9]. The 95% PI was computed by ±2 SEM for each band of WRS as well as the whole range of WRS and then upper and lower limits of the 95% PI was obtained for each score of WRS as a function of the number of test items.

Results

Test-retest reliability for the whole range of WRS

The data of test-retest results of the whole range of WRS with respect to 50 test words, the first 25 words and the first 10 words in each list were displayed for all subjects as a scattergram (Fig. 1). The range of presentation levels of test words were between 0 and 30 dB HL for all test conditions. Their means, SDs, correlations, SEs, SEMs, 95% CIs and 95% PIs were demonstrated in Table 2 for each number of test words. For 50 test words, Pearson coefficient of the correlation was 0.88 which is statistically significant at 0.01 level. The mean of WRSs at test was 64.57% with the SD of 23.61 and the mean at retest was 66.60% with the SD of 22.78 showing the mean of differences in WRS between the two tests (Md) as -2.03 with the SD of the differences (SDd) of 11.20. The 95% CI was ±0.92 and the 95% of PI was ±15.84. The one-way ANOVA revealed that there was a significant difference (p=0.000) among the results of each number of the test words. Post hoc tests demonstrated that there was a significant difference (p=0.000) between 10 and 25 words and also between 10 and 50 words; however, the difference was not significant (p>0.05) between 25 and 50 words.

For the first 25 words in each list, Pearson correlation coefficient was 0.76 statistically significant at the level of 0.01. The mean at the first WRS testing was 72.19 with the SD of 17.85 and the mean at retest was 74.05 with the SD of 17.61 showing the Md as -1.86 with the SDd of 12.35. Their 95%

Table 2. Means, standard deviations, post hoc test results, correlations, SE, SEM, 95% CI, and 95% PI of WRS tested by KS-MWL-A as a function of the number of test words

No. of test words	M1 (SD1)	M2 (SD2)	Md (SDd)	Post hoc test	r	SE (SDd/√n)	SEM (SDd/√2)	95% CI (±2 SE)	95% PI (±2 SEM)
50	64.57 (23.61)	66.60 (22.78)	-2.03 (11.20)		0.88 p=0.00	0.46	7.92	±0.92	±15.84
				p>0.05					
25	72.19 (17.85)	74.05 (17.61)	-1.86 (12.35)		0.76 p=0.00	0.69	8.73	±1.38	±17.46
				p=0.00					
10	71.98 (20.50)	73.43 (21.27)	-1.45 (18.54)		0.61 p=0.00	1.09	13.11	±2.18	±26.22

M1: mean of WRSs at test, M2: mean of WRSs at retest, Md: mean of differences between WRSs at test and retest, SD1: standard deviation of WRSs at test, SD2: standard deviation of WRSs at retest, SDd: standard deviation of differences between WRSs at test and retest, SE: standard errors of mean, SEM: standard errors of measurement, CI: confidence intervals, PI: prediction intervals, WRS: word recognition score, KS-MWL-A: Korean standard monosyllabic word lists for adults

CI was ±1.38 and the 95% of PI was ±17.46.

For the first 10 words in each list, correlation coefficient was 0.61 which is also statistically significant at the level of 0.01. The mean at test was 71.98 with the SD of 20.50 and the mean at retest was 73.43 with the SD of 21.27 showing the Md as -1.45 with the SDd of 18.54. The results also demonstrated the 95% CI of ±2.18 and the 95% PI of ±26.22.

Test-retest reliability for each score of WRS

Means and SDs of differences in WRS between test and retest, SEMs, and 95% PIs for the differences were described with respect to each band of WRS at test when using all 50 words at each list in Table 3. For the WRS band of 46−55%, the data showed the difference mean -4.32 with the SD 12.57, and the SEM 8.89 with the 95% PI ±17.78. As the WRS band increased up to the band of 86−100%, the SD decreased from 12.57 to 7.39. The data for the first 25 and 10 test words also showed similar trends to those for 50 test words as seen in Table 3. Based on these data, upper and lower limits of the 95% PI were calculated for each score of WRS from 0 to 100% as a function of the number of test items to be easily utilized in the clinics (Table 4). Values within the PI are not significantly different from the value shown in the WRS column (p>0.05)

Table 3. Means, standard deviations, SEM and 95% PI for each band of WRS tested by KS-MWL-A as a function of the number of test words

WRS band	No. of test words	Md	SDd	SEM (SDd/√2)	95% PI (±2 SEM)
46−55	50	-4.32	12.57	8.89	±17.78
	25	-7.85	12.77	9.03	±18.06
	10	-1.94	22.57	15.96	±31.92
56−65	50	-3.12	11.41	8.07	±16.14
	25	-3.43	12.68	8.97	±17.94
	10	-4.72	19.28	13.63	±27.26
66−75	50	-2.47	10.99	7.77	±15.54
	25	-2.45	11.91	8.42	±16.84
	10	-4.67	18.17	12.85	±25.70
76−85	50	-0.38	8.69	6.14	±12.28
	25	-0.58	11.93	8.44	±16.88
	10	2.12	16.13	11.41	±22.82
86−100	50	2.92	7.40	5.23	±10.46
	25	2.73	8.00	5.66	±11.32
	10	5.52	12.01	8.49	±16.98

Md: mean of differences between WRSs at test and retest, SDd: standard deviation of the differences between WRSs at test and retest, SEM: standard error of measurement, PI: prediction intervals, WRS: word recognition score, KS-MWL-A: Korean standard monosyllabic word lists for adults

Discussion

In this study, we tried to establish the test-retest reliability of KS-MWL-A regarding each score of WRS as well as the whole range of WRS as a function of the number of test words. Results of the whole range of WRS indicated that the test-retest reliability was high based on the high correlations and narrow CIs for 25 and 50 test words. As expected, the retest reliability of WRS for 10 test words was low, compared to the 25 and 50 test words. Previous studies [3,12] also reported that correlation became higher and SD was getting smaller and the CI was getting narrower as the number of test words increased in WRS testing. Both this study and aforementioned researches would recommend 25 or more test words for obtaining a reliable WRS.

As the presentation level increased from 0 to 30 dB HL, means of WRSs increased both at test and retest; however, the variation of differences between WRSs at test and retest became smaller, probably because of the ceiling effect toward the extreme band of 86−100%. These results are also consistent with the previous studies [3,4,6,11]. Correlation coefficients of this study are higher and CIs are narrower than

Table 4. Upper and lower limits of the 95% PI for each WRS tested by KS-MWL-A as a function of the number of test words

Score (%)	No. test words		
	50	25	10
0	0−10	0−80	0−10
2	0−12		
4	0−14	0−12	
6	0−16		
8	0−18	0−16	
10	0−20		0−20
12	2−22	4−24	
14	4−24		
16	4−28	4−32	
18	6−30		
20	8−32	4−36	0−40
22	10−34		
24	12−36	8−40	
26	12−40		
28	14−42	12−44	
30	16−44		10−50
32	18−46	16−48	
34	20−48		
36	20−52	20−52	
38	22−54		
40	24−56	24−56	20−60
42	26−58		
44	28−60	28−60	
46	30−62		
48	32−64	32−64	
50	34−66		20−80
52	36−68	36−68	
54	38−70		
56	40−72	40−72	
58	42−74		
60	44−76	44−76	40−80
62	46−78		
64	48−80	48−80	
66	52−80		
68	54−82	42−84	
70	56−84		50−90
72	58−86	56−88	
74	60−88		
76	64−88	60−92	
78	66−90		
80	68−92	64−96	60−100
82	70−94		
84	72−96	68−96	
86	76−96		
88	78−98	80−96	
90	80−100		80−100
92	88−100	84−100	
94	84−100		
96	86−100	88−100	
98	88−100		
100	90−100	92−100	90−100

PI: prediction intervals, WRS: word recognition score, KS-MWL-A: Korean standard monosyllabic word lists for adults

You and Lee [3] results for all test conditions, however. This is considered mainly due to the large group of subjects and their homogeneity in age in this study. As seen in Table 2, 95% PIs for the whole range of WRS are wider than 95% CI, which suggests that individual variance is greater than group variance. These results are also in consistent with the previous studies. In both large group and small group studies, PIs were reduced as the number of test words increased, which suggests that further analysis of PI for each score of WRS be needed for clinical utilization.

The whole range of WRS can be divided by 9 bands which consist of 0−14%, 15−24%, 25−34%, 35−44%, 45−55%, 56−65%, 66−75%, 76−85%, and 86−100%, so that the band of 45−55% is positioned at the center band. In this study, as expected, the SD of differences between WRSs at test and retest was largest at the center band and gradually decreased as the band level went up to the highest level for all three conditions of the number of test items. It can be theoretically inferred regarding the normal distribution that if data were obtained at WRS bands lower than the center band, SDs at lower bands would be also smaller than that at the center as SDs at upper bands were. That is, the variances of upper bands of 86−100%, 76−85%, 66−75%, and 56−65% would be equal or at least similar to the lower bands of 0−15%, 16−25%, 26−35%, and 36−45%, respectively. Thus, it can also be inferred that as WRS band increases, 95% PI of each band also decreases as SD does, because PI is calculated by the SEM which is directly affected by SD.

In this study, the intra-subject variability in WRS is described by the ±2 SEM for 95% PI in Table 2 and 3 as recommended by previous researches [3,8,9]. The SEM is different from the SE which refers to the SD of sample means as explained earlier. The SEM is directly related to the reliability of a test with respect to an individual performance, that is, the wider the PI, the lower the reliability of the test. Thus it can also be asserted that the more the number of test words, the higher the reliability of the test. However, testing time is also an important factor regarding clinical efficiency. That is why it is valuable to generate the table showing the upper and lower limits of 95% PI as a function of the number of test items, which can be easily used at clinical settings when interpreting individual retest results. If a difference between test and retest WRS score is greater than double of the SEM, then it means a statistically significant variation with respect to the 95% PI. The upper and lower limits of the 95% PIs for each score of WRS in this study show similar trends to those of 95% critical differences about English WRS for adults reported by Thornton and Raffin [6], although they calculated the 95% critical differences based on the binomial

confidence intervals.

As aforementioned, PIs are affected by the number of test words as well as the WRS band level as seen in Table 3. For example, if WRS measured by using 25 test words was 60% before auditory training, the upper limit of the PI of this condition would be 76% as seen in Table 4. Thus, WRS of 80% or greater be interpreted as a significant improvement after training. If the 50 test words were used, then the upper limit of the PI would be 76%. Thus, 78% or greater at retest would be accepted as a significant improvement. For the 10 test words, however, the upper limit of the PI would be 80%, thus only 90% or 100% at retest would be accepted as a significant improvement. In the other example, if WRS for 50 test words was 30% without fitting hearing aids, the upper limit of the PI of this condition would be 44% as seen in Table 4. Thus, the WRS of 46% or greater be interpreted as a significant improvement after fitting the hearing aids. If the 10 test words were used, then the upper limit of the PI would be 50%. Thus, 60% or greater at retest would be accepted as a considerable improvement. In sum, it would be important to apply the PI values as a function of the number of the test words in Table 4 for interpreting individual retest results.

Conclusion

This study aimed to investigate the test-retest reliability of WRS testing as a function of the number of test words. Twenty-five or greater test words are recommended for reliable WRS measurement for adults, based on higher correlations, narrower CIs and PIs compared to those of 10 test words. When interpreting retest results, 95% CI for the whole range of WRS for each number of test words would be useful for group data. For individual data, however, 95% PI at each score of WRS for each number of test words would be more useful. If WRS testing with 10 test words is necessary for some individuals for some reasons, then 95% PI for 10 test words should be applied for interpreting retest results of that individual.

Acknowledgments

This research was sponsored by a grant from the Korean Ministry of Trade, Industry & Energy (Project 10041529).

REFERENCES

1) Kim JS, Lim DH, Hong HN, Shin HW, Lee KD, Hong BN, et al. Development of Korean standard monosyllabic word lists for adults (KS-MWL-A). Audiology 2008;4:126-40.
2) Korean Agency for Technology and Standards. Acoustics-Audiometric test methods-Part 3:speech audiometry. KSI ISO 8253-3. Seoul: KATS;2009.
3) Yoo BM, Lee JH. Prediction interval of word recognition score using Korean standard monosyllabic word lists for adults (KS-MWL-A). Audiology 2014;10:35-42.
4) Lee HW, Lee KW. The test-retest reliability of the word list of Korean speech audiometry for preschoolers. Audiology 2014;10:25-34.
5) Yoon JY, Lee JH. The test-retest reliability of Korean standard language lists for schoolchildren in speech audiometry. Audiology 2015; 11:26-36.
6) Thornton AR, Raffin MJ. Speech-discrimination scores modeled as a binomial variable. J Speech Hear Res 1978;21:507-18.
7) Demorest ME, Walden BE. Psychometric principles in the selection, interpretation, and evaluation of communication self-assessment inventories. J Speech Hear Disord 1984;49:226-40.
8) Hopkins WG. Measures of reliability in sports medicine and science. Sports Med 2000;30:1-15.
9) D'Haenens W, Vinck BM, De Vel E, Maes L, Bockstael A, Keppler H, et al. Auditory steady-state responses in normal hearing adults: a test-retest reliability study. Int J Audiol 2008;47:489-98.
10) Kim SR, Lee J. Test-Retest Reliability of Bone-Conducted Auditory Steady-State Response. Audiol 2010;6:50-4.
11) Grange ME. Test-retest Reliability in word recognition testing in subjects with varying levels of hearing loss [dissertation]. Provo, UT: Brigham Young Univ.;2013.
12) Hong SA. Test-retest reliability of Speech Discrimination Test using the monosyllabic word lists. Korean J Audiol 2002;6:128-35.
13) Kim AK. The test-retest of the monosyllabic word lists on word recognition measurement in normal hearing adults [Master's thesis]. Department of Audiology; Hallym Univ. of Graduate Studies;2008.
14) International Organization for Standardization. Acoustics-Audiometric test methods-Part 3: speech audiometry. ISO 8253-3. Geneva: ISO;2012. p.1-36.
15) Cho SJ, Lim DH, Lee KY, Han HK, Lee JH. Development of Korean standard bisyllabic word list for adults used in speech recognition threshold test. Audiology 2008;4:28-36.