

20 MOST FREQUENTLY ASKED QUESTIONS REGARDING RECORDED WORD RECOGNITION TESTING

WRS = word recognition score, WR = word recognition, SAS recordings = South African Speech Audiometry recordings, SNR = signal to noise ratio, PI = Performance-Intensity Curve, APD = Auditory Processing Deficit

1. ***What should I NOT be doing when conducting WR testing?***

Things **NOT** to do (Hornsby & Mueller, 2013):

- Perform WR testing using live voice testing. This will provide unreliable results.
- If a single presentation level such as +30/40 dB SL is used, it will result in patient's maximum WRS.
- Only testing in quiet.
- Making random guesses when interpreting whether two scores differ significantly from each other.
- Conducting only a section of the wordlists, for example presenting 15 words from the FVEWA and not 25 words (full list).
- Using the term 'speech discrimination testing' in reports and during feedback with patient.

2. ***Why should I use the term "word recognition" and not "speech discrimination" be used?***

The incorrect term *speech discrimination* was used clinically and in research literature up to the 1990s (Hall, 2017). Hearing and repeating a word are, in fact, a *recognition* task. Can you remember the auditory hierarchy as described by Norm Erber (1982). Hearing a word and repeating it would be classified on the *identification* level whereas *discrimination* is on a different level where two stimuli are often provided and the patient must say whether they are different or the same. Please use the correct term *recognition* when you are thinking, reasoning, speaking and writing reports. It is also important to train your ENT regarding the use of the correct term.

3. What *SHOULD* I be doing when conducting WR testing?

Things to **do**:

- Always use recorded materials (Mueller & Hornsby, 2020) and try to make recorded materials your goal even when testing a child (Madell, 2011). Many variables are excluded when using recorded materials, and results are reliable and can be successfully compared across sessions.
- Have a sip of tea/coffee whilst presenting recorded materials!
- Choose a presentation level to maximize audibility without causing loudness discomfort.
- Refer to the binomial model of Arthur Boothroyd (2019) to determine when any two scores are significant different from each other.
- Use the Dubno et al (1995) norms on the SRINT CHART (50 words) when using the NU6-II to determine when results are “normal” compared to the PTA3.
- Use the Dubno et al (1995) norms on the SRINT CHART (25 words) when using the NU6-I (half lists) to determine when results are “normal” compared to the PTA3.
- Conduct phoneme scoring when using the Arthur Boothroyd (AB) lists.
- Perform word recognition in quiet and with competing babble noise as part of the **basic test battery**.

4. How should I set the VU meter to 0 when playing the calibration track before conducting recorded WR testing?

On the audiometer the VU meter knob/setting for either the internal recording or an external device (tablet or smartphone) is different as the VU meter knob/setting used for live voice testing such as for speech recognition testing. When using an external device, ensure that its volume is on maximum. For most audiometers once of calibration of the VU meter is required unless you use different external devices you must calibrate that device. Present the calibration track. You will hear a continuous pure tone @ 1000 Hz at 70 dBHL) and set the VU meter on 0.

You are ready to fly....



5. Despite calibrating the VU meter, I still notice that the VU meter peaks either below or above 0 when presenting the recorded word lists. Does this mean the recording is either too soft or too loud?

No not at all. When using live voice (which we may only apply when conducting SRT testing) we aim for the VU meter to peak at 0, which means the highest intensity of a sound in the word is peaking at 0. For several years now in the recording world, material is recorded applying **averaging volume**. All the SAS recordings have been technically recorded using averaging volume, meaning the “whole” word (and thus all the speech sounds) have been averaged out. You don’t have to be concerned that the words being presented are either too loud or too soft. Just ensure that you have calibrated your VU meter.

6. I am aware that WR testing is a supra-threshold measurement, but at which intensity should I present my first word list in quiet?

Supra-threshold measurement defines that the intensity level/s at which the word lists are presented are well **ABOVE** threshold results (such as PTA and SRT). “Well above” refers to words being audible for the patient despite his hearing loss.

The goals and purpose of performing WR testing in your practice/clinic/hospital will determine the intensity levels at which you will present the recorded lists. When using procedure code 821110, SRT & WR testing should be done. WR in quiet should be done 2 intensities per ear thus two wordlists.

When testing adults with mild to moderate sensory-neural sloping/steep sloping or severe losses, the 2000 Hz sensation level (SL) by Guthrie and Mackersie (2009) are recommended. This method should yield the maximum WR score. The second point on your performance-intensity curve can either be at a louder intensity or at a softer intensity (for example 50 dB which correlates with normal conversation speech).

When testing children, I would like to refer you to an article by Jane Madell on AudiologyOnline (2011).

For adults you should always ask yourself “**How good does it get for this patient?**” Can he/she perceive words correctly when they are audible? (Mueller & Hornsby, 2013). You may not penalise a patient for not obtaining a good WR score when the words were not audible to the patient. The 2 kHz sensation level method ensures that words are audible for patients with a mild to moderate sensorineural sloping/steeply sloping hearing loss and/or severe hearing loss (Guthrie & Mackersie, 2009)

According to Mueller (2016) you have three choices: a relatively high level (80/90dBHL), the level of average speech (50dBHL) and a softer (35 dBHL) than average speech level.

Janet Schoeplin reports that patients with **normal hearing** sensitivity should obtain a maximum WR score 25 – 40 dBSL. Thus, if the PTA or SRT is 10 dB, you may expect to obtain 100 % WRscore at either 35, 40, 45 or 50 dB. Testing and expecting 100 % WR in quiet at 30 dB is not the Rule and 30 dB represents very soft speech. It is recommended to present the first list at 40 dBHL. If only one point/intensity is done per ear, the **screening code for Speech Audiometry** can be charged (821115)

For hearing impaired children, the purpose of using speech perception tests is not to have the child obtain the best possible score (Madell, 2011). The purpose is

- to monitor the child's technology.
- to determine if the amplification settings need to be modified.
- to determine areas of need for habilitation/rehabilitation.
- to plan for educational placement.

The following test protocol is recommended by Madell (2011). Test at:

- a normal conversational level (50 dB HL) in both ears and/or free field;
- a soft conversational level (35 dB HL) in both ears and/or free field;
- a normal conversational level in competing noise (50 dB HL with +5 dB SNR) in free field/sound field.

7. How many points (different dB levels) should I have on my performance intensity (PI) curve?

Again, it will depend on your setup, goals and purpose of doing WR testing. Arthur Boothroyd recommends the use of 3 points on the incline and 3 points on the plateau. This is easily performed with a 10-word list like the AB. McCreery (2010) recommends the use of 3 or 4 points per ear when using the AB wordlists. Should the procedure code 821110 be used, SRT and two point per ear on the PI-curve should be conducted.

Should you strive to determine whether the patient has a roll-over, you should present the words at high intensity levels, and you absolutely must use recorded material. First, the maximum WRS must be obtained. Thereafter the minimum WRS (which should be at an intensity level above the maximum score) should be obtained. This will usually be at 80/90/100 dBHL. You should stop testing if the patient's UCL for speech occurs. Determine the roll-over index or ratio by using the following formula:

$(PB \text{ max} - PB \text{ min}) \text{ divided by } PB \text{ max} \geq 0.4 / 0.45 - \text{positive roll-over}.$

Keep in mind only 8% of 231 patients with confirmed unilateral acoustic neuromas had significant roll-over effect (van Dijk et al., 2000).

Van Dijk et al (2000) concluded that the roll-over is not characteristic for patients with an acoustic neuroma. They found only 3% with Roll over Index (RI) exceeding 0.45 and 5% of patients with $RI \geq 0.42$.

Many roll-overs are found among patients with cochlear hearing losses and words at loud intensities either distort or are just too loud to perceive.

8. *Just refresh my memory - what are the reasons for performing WR testing in quiet as part of the audiometric test battery?*

WR testing was first performed in the USA after WWII, since an audiogram is not sensitive enough to predict which speech sounds are audible and which are not.

- This test allows you to determine how a person's loss of hearing sensitivity affects his/her ability to perceive speech sounds when speech is soft/normal/loud.
- This is a supra-threshold measurement, so speech is made audible again for the patient. Now you can determine if a patient's WR skills are excellent or poor.
- You can determine if there is a significant difference in WRS between the two ears, especially when the audiograms are symmetrical. This is a red flag finding.
- You can determine if a significant change in WRS has occurred since the previous test (monitor over time).
- You can determine whether a patient is better off with a hearing aid than without one. (Aided vs unaided score comparisons).
- You can determine if a patient's WRS agrees with his PTA loss (possible with the NU 6 II & NU 6 I results).
- You can apply results during counselling of the patient and his loved ones

9. *I notice that differentiating between cochlear and retrocochlear pathologies is not listed as one of the reasons for performing WR testing?*

That is correct! A subjective test such as WR in quiet is the poorest test available for distinguishing cochlear from retrocochlear site-of-lesion (Turner et al., 1984). The roll-over effect is often seen amongst patients with cochlear hearing losses (Harner et al, 2000). A significant roll-over is seldom found amongst cases with confirmed acoustic neuromas (van Dijk et al, 2000)

REMEMBER: You also cannot diagnose/report a roll-over if you have not used recorded WR materials.

10. So, if WR does not add diagnostic value to your test battery, how can we apply it to be useful?

WR in quiet does not add diagnostic value to your test battery (Hall, 2017), so start using WR with a **competing signal**. The multi-talker babble noise on our recordings consists of 12 talkers and are described as being “energetic” masking/noise – you cannot determine what the people are saying in the background such as the case is at a party/restaurant or large gathering. Most recordings are done with a fixed +5 and +10 dB SNR. Some wordlists are recorded as adaptive recordings (using Channel 2).

4-talker English & Afrikaans Adult and children’s babble are also available on some recordings. 4-talker babble is described as “informative noise” and requires peripheral and central auditory skills to recognize the word/sentence.

Determine if your patient’s WR score differs significantly when testing using a +5 dB and +10 dB SNR compared to when testing in quiet. Refer to the Binomial model to determine whether the two scores differ significantly from each other or not. Test conditions need to be the same for comparison reasons either both in free field/sound field OR under headphones/inserts.

Testing WR in noise can help you to better identify auditory neuropathy cases or hidden hearing losses or APD conditions amongst adults.

11. WR testing in the presence of noise, is recommended to be part of the basic test battery. How do I do this?

The “noise” cannot be speech or white noise (as used during masking). Noise should be “speech like” thus either 8 – 12 people talking which is called multi-talker noise OR a 4-talker babble noise.

As soon as the maximum score is obtained in quiet, it is recommended testing words with competing babble noise in each ear separately. This can be done at the dB level where the patient **with hearing loss** obtained his/her maximum score in quiet. For a patient with **normal hearing sensitivity** at normal conversational speech thus (50 dBHL).

Jane Madell recommends testing children at 50 dB using +5 dB SNR since this represents a classroom environment. Should your patient's score be significantly worse than that in quiet (refer to the binomial model) then you may have identified “hidden hearing loss”, auditory neuropathy, auditory processing difficulties, etc depending on the case history/other results.

Present a different word list with babble noise via headphones/inserts in each ear. For patients with normal hearing, mild, moderate degree of hearing loss a +5 dB SNR can be used. For patients with a moderate-severe to severe degree hearing loss a

+10 dB SNR can be used. Both SNR's can also be used for comparison. The **CAPD procedure code 821400** (specified as speech-in-noise testing) can be used.

12. Can I do WR in noise with children as part of the APD test battery?

Yes, words in noise testing are a must for children with possible APD. All the SAS recordings are recorded with multi-talker babble noise and or classroom noise. Jane Madell (2011) recommends three test conditions for children suspected of having APD:

- *50 dB HL with +5 dB SNR
- 50 dB HL with 0 dB SNR
- 35 dB HL with 0 dB SNR

She found that 187 children with normal hearing scored 88% or better in all three of the above-mentioned conditions. A significantly lower score indicates that the child is functioning below normal limits and has difficulty hearing in noise.

The CAPD procedure code 821400 (specified as speech-in-noise testing) can be used.

**The first option is available on our recordings.*

13. What about WR lists in other languages?

A standardized isiZulu wordlist has been developed at the University of KZN. Currently a Xhosa wordlist is being developed at UCT. Other African languages have been recorded are not standardised on the SA population.

Watch this space for future recordings by SAS in 2026. Most African languages' words consist of two or more syllables which would be easier to recognise due to lexical and contextual clues.

14. What about materials for children? Can I still use monitored live voice when testing children?

SAS has developed CVC open-set word lists in quiet and in noise in English and Afrikaans. It consists of 10 lists with 10 words in each list and no consonant clusters occur. The vowels are isophonemic. There are two types of noises: classroom children's babble noise and competing 4-talker babble noise. Phoneme scoring is recommended.

Lida Müller has granted SAS permission to record the ODPI (Ouditiewe Diskriminasie Prent Identifikasie Toets) and the English version, the WIPI. These are closed-set word lists. Testing children is a challenge and using recorded materials should be the goal but may not always be possible.

15. How do I describe my test results when WR testing was performed as part of the basic test battery?

Remember to mention in your report that you have used **recorded word lists** and supply the **name** of the materials used.

For children, refer to Madell et al. (2011) categories for scoring Speech Perception scores.

WRS (% correct)	Word recognition Ability
90 - 100	Excellent
80 – 89	Good
70 – 79	Fair
< 70	Poor

Also see the classification system as described by Kramer & Brown (2019), commonly used for adults.

WRS (% correct)	Degree of Impairment	*Word Recognition Ability
90 – 100	None	Excellent/Normal
75 – 89	Slight	Good
60 – 74	Moderate	Fair
50 – 59	Poor	Poor
< 50	Very Poor	Very Poor

*It is advisable to refer to WR ability in you report

16. What WR in quiet results are expected from different type of hearing losses?

Patients with normally hearing sensitivity are expected to obtain scores between 90 – 100 % being described as excellent.

Conductive hearing losses should obtain 80 – 100% WR scores.

Sensory-neural hearing losses can obtain scores anything between 0 – 100 % WR in quiet. (Bess, in Gelfand 2016)

17. When should I be concerned of a particular WR test score?

For children who make use of amplification, Madell is concerned when a score is below 70% despite having good technology and good real-ear measurements.

Some patients may perform **poorly or very poorly** (60% or lower) despite the words being presented at an audible level and this may be a reason for concern, especially if it is not expected when the audiogram is taken into consideration.

When patients obtain significantly poorer WRS in quiet and/or in babble noise in one ear compared to the other despite having **symmetrical hearing loss**, it is a cause for concern. Note that research has proven that the above phenomena is found if a patient is fitted with one hearing aid and the poorer WRS are found in the non-aided ear due to sound deprivation.

Patients who have **outer and inner hair cell cochlear damage** (hearing loss worse than 55 dBHL) may perceive words as too loud and a decrease in scores is observed as the intensity increases. This phenomenon is also observed in patients with “**dead zones**” and they may be described as the words being distorted.

Patients with **neural presbycusis** have an inability to sustain neural activity when the auditory nerve is “stressed” at high intensity levels and may show a roll-over (Kramer & Brown, 2019). Patients with **APD and hearing loss**, may also obtain poorer than expected results in both ears. Although the last four cases are not a medical concern, these variables will affect the choice and setting of amplification devices and rehabilitation.

18. How do I know if any TWO are SIGNIFICANTLY different from each other or not?

Refer you to the Binomial model of Arthur Boothroyd (with permission to distribute). Remember ANY two scores can be compared with each other to determine whether the two scores differ from each other or not. For example:

- Is the WR score in quiet of the **right ear** different from the WR in quiet score in **left ear** despite having a symmetrical audiogram.
- Is the WR score in **quiet** of the right ear different from the WR score with babble **noise** in the right ear different from each other.
- Is the WR score before and after medical treatment or surgery different from each other.
- Is the WR score without and with amplification different from each other.

Herewith a short explanation how to use the chart. Use the lowest of the two scores first. Find the score below “lower score” section (far left row). Intersects with the

number of items of the TEST scored (column). There are four columns. 25 items, 50 items, 63 and 100 items. The first two columns will apply for the SAS recordings.

- Should AB word lists (phoneme scoring), FVEWA (whole word scoring), CVC & KVK (phoneme scoring) wordlists be used, the **25 item column** will be used to determine the range of significance
- The **50 item column** will be used when the NU6-II (ordered by difficulty) wordlist has been used. Even if you only presented the 10 most difficult words, it remains a 50 item wordlist.

Example 1 (see a section of the table below): The right ear's maximum score is 64% and the left ear's maximum score is 92%. Are the scores significantly different from each other? The FVEWA 25 item word list was used. The poorer score of the two is 64% (left ear). Search for 64% in the row under the "lower of the two scores". Then trace your finger horizontally to the number in the second column (25 items) which is 88. The range is 64 to 88%. The left ear's score is 92%. Does 92% fall within the **range**? 92 % falls **OUTSIDE** the above range, which means the two scores **are significantly different from each other**. Should the second score fall **inside the range**, the two ears/scores are **NOT significantly** different from each other.

Example 2 (see a section of the table below): WRS are 52% and 64%. The AB lists were used, using phoneme scoring, thus the AB is a 25 item wordlist (10 x 2.5 = 25) and the 25 column will be referred too. The lower score of 52% must be found in the left column. Horizontally move your finger across to 25, which will be 79%. The range for not being significant is 52 to 80%. Your second score is 64%. This score is within the range and thus does **not significantly different** from the other. Should the second score be above 80%, it would have fallen **outside the range** and would thus have been **significantly different**.

Lower Score (%)	Number of test items			
	25	50	75	100
	Significantly (p<.05) higher score (%)			
50	76	70	66	64
51	80	72	67	65
52	80	72	68	66
53	80	72	69	67
54	80	74	70	68
55	80	74	71	69
56	84	76	72	70
57	84	76	73	71
58	84	78	74	72
59	84	78	75	73
60	84	80	76	74
61	88	80	76	75
62	88	80	77	76
63	88	82	78	76
64	88	82	79	77

19. Why is the above calculation necessary?

As stated previously, you should not guess whether two scores are significantly different from each other. Your findings should be based on data that has been studied to determine a statistically significant difference. For medico-legal purposes, this is evidence-based practice, and you are not left guessing whether results are significantly different or not. You can also prove to the patient, significant other or medical schemes that the patient is performing significantly better with his amplification device than without it.

20. Any tips when doing recorded WR testing?

Just do it! Ask for help to get set up. It is not difficult; it only requires a few changes. No audiometer is too old or too new to use with recorded materials. The older version audiometers will have a CD/TAPE button and a connector button at the back of the audiometer. Newer models will have an Ext A/B (external) button and plug. You can download the sound files onto a tablet/smartphone and plug it into the audiometer via an audio cable. On computerized audiometers, word lists can be loaded internally by yourself or the technician.

It is not recommended to use pre-recorded speech materials from other countries since the presenter, or the babble noise present with an accent. Research have shown that other accents affect the results negatively.

Read your manuals as provided. Most wordlists have copyright and may not be distributed. Ask questions in you are not sure! This has been a learning curve for us all, and we are still learning as time passes and as research is being performed. Good luck!

References

- Boothroyd, A: Binomial model. Personal communication with Roets & Hornby
- Dubno, JR, Lee FS, Klein AJ, Matthews LJ & Lam, CF. (1995). Confidence limits for maximum word-recognition scores. *Journal of Speech & Hearing Research*. Volume 38, 490 – 503, April
- Erber, ND (1982). Auditory training. Washing DC: AC Bell Association for the Deaf.
- Gelfand, SA (2016). *Essentials of Audiology*. Fourth Edition. Thieme Publishers.
- Guthrie LA & Mackersie CL. A Comparison of Presentation Levels to Maximize Word Recognition Scores. *Journal of American Academy of Audiology* 20:386-390, 2009
- Hall. W, James III (2017). Rethinking the Audiology Test Battery. *Audiologyonline*
- Harner, SG., Fabry, DA & Beatty CW. (2000). Audiometric Findings in Patients with Acoustic Neuroma. *The American Journal of Otology*, 21:405-411.
- Hornsby, B & Mueller, HG. Monosyllabic Word Testing: Five Simple Steps to Improve Accuracy and Efficiency. *AudiologyOnline*, Article#11978. Retrieved from <http://www.audiologyonline.com/>. July, 2013
- Kramer, S & Brown, D.K (2019). *Audiology: Science to Practice*. 3rd Ed.
- Madell, Jane. (2011) Pediatric Amplification: Using Speech Perception to Achieve Best Outcomes, Retrieved from www.audiologyonline.com February.
- Madell, JR., Klemp, EJ., Batheja, R & Hoffman, R. (2011). Evaluating Speech Perception Performance. *Audiology Today* Sept/Oct 52 – 57.
- McCreery, R. et al (2010). Performance-intensity function for normal-hearing adults and children using the CASPA, *Ear & Hearing*.
- Mueller, HG. (2016). Signia Expert Series: Speech-in-Noise Testing for Selection and Fitting of hearing Aids: Worth the Effort? Retrieved from www.audiologyonline.com Sept.
- Mueller HG & Hornsby BWY (2020). 20 Q Word Recognition Testing: Let's just agree to do it right!. Course 34293. Retrieved from www.audiologyonline.com Feb.
- Turner RG., Shepard, NT & Frazer GJ. (1984). Clinical Performance of Audiological and Related Diagnostic Tests. *Ear and Hearing*, Vol 5, no 4 p187 – 194.
- Schoepflin, JR (2012). Back to Basic: Speech Audiometry. Retrieved from www.audiologyonline.com July

Van Dijk JE, Duijndam, J & Graamans, (2000). Acoustic Neuroma: Deterioration of Speech Discrimination related to thresholds of pure-tone Audiometry. *Acta Otolaryngology*, 120: 627 – 632.

Compiled by Rene Hornby & Rozelle Roets, June 2025