

# **Supplemental Material S7. Considerations for interrater reliability: Training materials.**

Score 0	Response/Target	Score 1	Response/Target
More than one articulatory error or distortion	/tisʌm/ (teasom)	Minor consonant errors (no more than one feature from the target)	/zisʌn/ (zeason)
Omission of a sound/syllable	/sisʌ/ (seas-uh)	Excessive lengthening	/si:sʌn/ (sea:son)
Addition of a sound/syllable	/sisʌnʌ/ (season-uh)	Incorrect stress pattern or syllable segregation	/siSʌN/ (seaSON)
Vowel substitution	/saisʌn/ (sigh-son)	Vowel distortion (sounds like a poor exemplar of the target)	/sisʌn/ (i~l)
Consonant error (more than one feature from the target)	/sidʌn/ (seadon)		
Vowel or consonant error with prosodic error	/sid: ʌn/ (sea: don)		
Unintelligible production			
Greater than 1 prosodic error	/s:i:sʌn/ (s:ea:son)		

1. The interrater completed the initial training on a practice video of the probe. This is compared immediately, and any variations between the PI and interrater will be discussed before further rating activities.
2. The PI can state the target word up to twice, and any of the participants' attempts to produce the words are scored.
  - a. If PI states the target word more than twice in error, then the score is after the second repetition.
3. Score the most accurate attempt by the participant.
4. Use the chart to identify errors, both prosodic and phonetic errors are identified.
5. Close distortions and phonetic substitutions (alone without other errors) were scored in the same way as a 1.
6. Use the chart to calculate total whole word accuracy for each production.
7. The rater will complete interrater reliability without knowledge of the time period of the probe or the sound being targeted during each treatment phase.

8. Prosodic elements are based on how the PI says the word, although the PI may make some adjustments given dialectal differences (for example, "vehicle" for participants 1 and 3). Accuracy can be judged based on the manner in which the model is presented.
9. If additional variations occur or the rater has a question during reliability assessment, the PI and interrater will discuss these differences and arrive at a consensus based on the materials above. This occurs without referencing a specific probe number to ensure no additional bias in the discussion.