A Quantitative Comparison of Articulation Assessments for Different Consonant Error Profiles

Courtney Seidel, M.S., CCC-SLP & Tristan Mahr, University of Wisconsin-Madison

BACKGROUND

- In Wisconsin, one way students qualify for speech services by scoring at least 1.75 standard deviations below the mean on a test of articulation or phonology.
- Standardized articulation tests differ in scoring procedures and in the number of opportunities the child has to produce a speech sound. A pervasive consonant error may be scored in one test but ignored by another test.
- A student referred for a phonological delay may therefore have their particular error profile systematically ignored by an assessment.
- **How we do know which test will be most sensitive for this particular student's error profile?**
- When we encounter a student with pervasive /r/ errors, which test will provide a thorough sample of the student's speech profile?
- The focus of our study is to compare multiple articulation assessments for sensitivity to later developing consonants.

METHODS

Participants
- Eight 6 to 8 year old students (3 boys, 5 girls, mean age = 7;5)
- Referred for articulation testing in Middleton–Cross Plains Area School District

Materials
- Eight students received the following assessments:
  - Goldman-Fristoe Test of Articulation (GFTA-2)
  - Khan-Lewis Phonological Analysis (KLPA-2)
  - Hodson Assessment of Phonological Patterns (HAPP-3)
- Two students also received the Bankson-Bernthal Test of Phonology (BBTOP).

Transcription Conventions
- All tested words were transcribed in IPA.
- Each consonant error was classified as a substitution, deletion or distortion.
- Distortions were defined as productions that cannot be transcribed without an IPA diacritic (e.g., intermediate and exaggerated productions).
- Consonantal /r/ and vocalic /ɚ/ ("er") were considered separate sounds.

Analysis Procedure
- Consonant errors were scored according to the protocol of each test.
- For each consonant error, we recorded whether the test included or excluded the error when determining the norm-referenced scores.
  - For example, vocalic /ɚ/ errors are always scored on the BBTOP and KLPA-2, just once on the GFTA-2 (*car*), and zero times on the HAPP-3.

Test Analyses
- For each test, we calculated the percentage of target words with scored instances of the sounds /s,z/, /r,ɚ/ and /l,r,ɚ/:

Class	BBTOP	GFTA-2	HAPP-3
s/z	16 (20%)	9 (17%)	22 (44%)
er/r	21 (26%)	9 (17%)	8 (16%)
l/er/r	39 (49%)	18 (34%)	18 (36%)

Predictions
- Students will be ranked differently within each test.
- Compared to the KLPA-2 and GFTA-2, the HAPP-3 will be more sensitive to /s,z/, dental errors (i.e., lisping).
- Compared to the HAPP-3 and GFTA-2, the KLPA-2 will be more sensitive to /l,r,ɚ/ errors.

RESULTS

Error Permissiveness
- Eight students produced 348 consonant errors.
  - 38% of errors were ignored by the four tests
- BBTOP was least permissive because any consonant error counts as a Word Inventory error.

Test	N	# Errors	Permitted
BBTOP	2	38	0%
KLPA2	8	173	31%
HAPP3	8	175	41%
GFTA2	8	173	52%

Error Profiles
- Students showed the following error profiles:

Subject	ɚ	r	s	z	l	Other	Sum	Profile
Lisa	17	17	0	0	24	15	73	l (33%), ɚ (23%), r (23%)
Maggie	18	12	0	0	0	1	31	ɚ (58%), r (39%)
Marge	0	0	22	15	0	0	37	s (59%), z (41%)
Milhouse	13	16	1	0	0	1	31	r (52%), ɚ (42%)
Nelson	13	17	23	16	3	10	82	s (28%), r (21%), z (20%)
Patty	6	2	0	0	0	9	17	ŋ (53%), ɚ (35%), r (12%)
Ralph	17	17	9	5	2	3	53	ɚ (32%), r (32%), s (17%)
Selma	7	8	0	0	0	9	24	ŋ (38%), r (33%), ɚ (29%)
(all)	**91**	**89**	**55**	**36**	**29**	**48**	**348**	ɚ (26%), r (26%), s (16%)

- The most frequent errors occurred on the liquids /r, ɚ,l/ and stridents /s,z/

Permissiveness of Error Profiles
- As predicted, HAPP-3 scored the greatest proportion of /s,z/ errors.
- As predicted, KLPA-2 scored the greatest proportion of /r, ɚ, l/ errors.

	r, ɚ, l	r, ɚ	s, z
GFTA2	47%	44%	49%
KLPA2	94%	94%	2%
HAPP3	55%	49%	80%

Concordance
- We tested the inter-test reliability of the GFTA-2, KLPA-2 and HAPP-3 using Kendall's coefficient of concordance.
- Students were ranked within each test by their percentile scores.
- There was not significant concordance among the three tests,
  - W(8 subjects, 3 judges) = 0.55, χ^2(7) = 11.6, p = 0.12.

Concordance among articulation tests
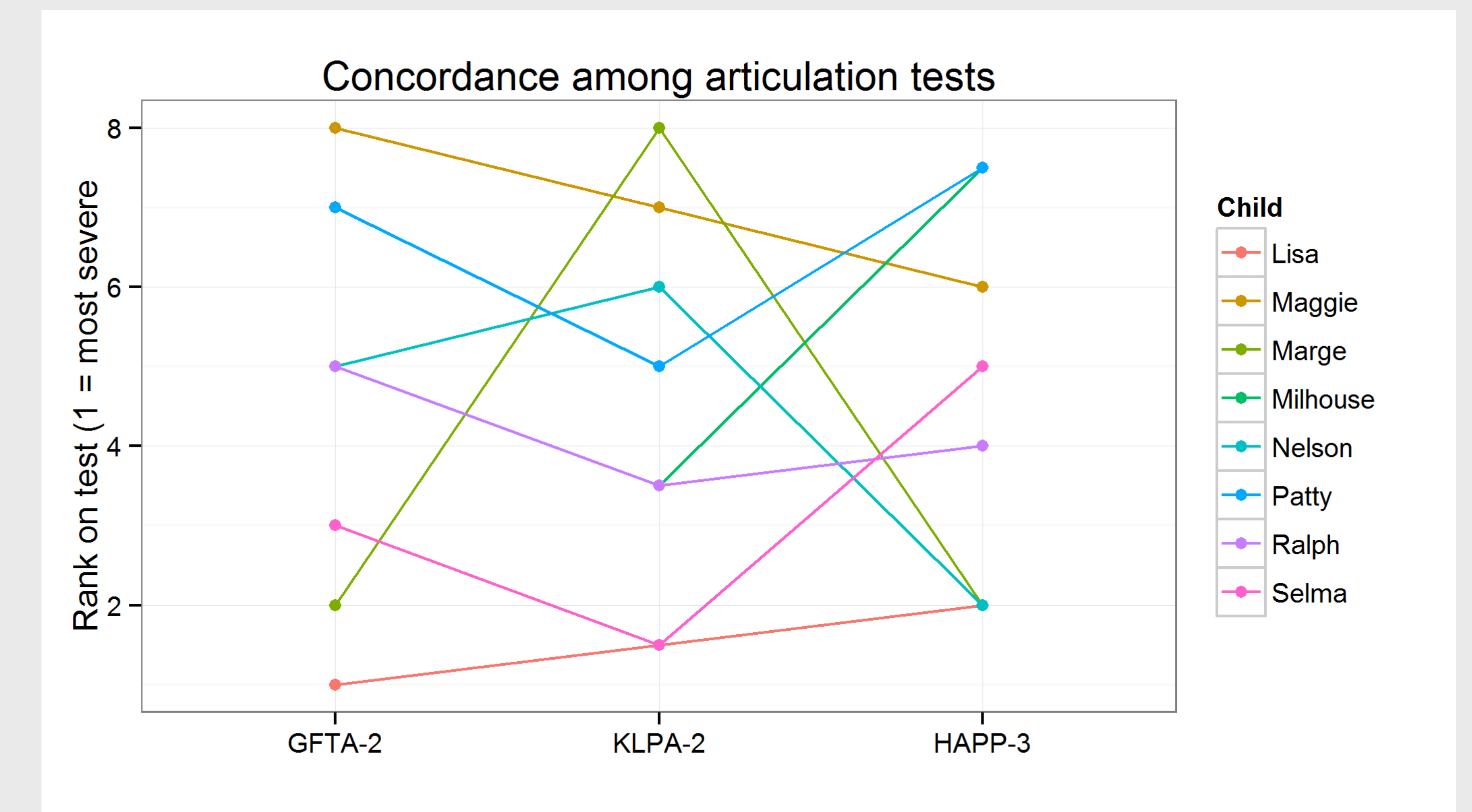
DISCUSSION

Which tests do I choose?
- Word-level testing, as on the BBTOP, maximizes sensitivity.
- Student with /r,ɚ/ errors (gliding, vowelization): BBTOP or KLPA-2
- Students with /l,r,ɚ/ errors (gliding): BBTOP or KLPA-2
- Student with /s,z/ errors (lisping): BBTOP or HAPP-3

Comments on Specific Tests
- GFTA-2 is designed to examine a student's sound-place consonant inventory. It provides few production opportunities (initial, medial, final for most consonants).
- KLPA-2 uses the same set of productions as a GFTA-2 administration; however, it scores them according to certain phonological processes. It provides many more scorable trials than the GFTA-2.
- HAPP-3 specifically looks at consonant omissions and specified substitutions (e.g., backing, fronting, gliding). It excludes distortions, approximations, and substitutions that are not phonological processes. It also excludes /l/ and /r/ errors in post-vocalic position (e.g., candle, fork).
- BBTOP provides an extensive number of scorable trials (especially for later-developing sounds) and offers three composite scores, including articulation and phonology.

Study Limitations
- We could only look at the Word Inventory score on the BBTOP because the Consonant and Phonological Process inventories translate errors into a 4-point scale.
- On the HAPP-3, we used the 7;11 norms for children 8 years and older in order to compare to results from all three tests. (Their actual percentiles would only be lower compared to age-matched peers.)
- Modest sample size

ACKNOWLEDGEMENTS