Text Messages: Examining Different Estimates of Text Complexity

Joseph F. T. Nese
Julie Alonzo
Gina Biancarosa
Akihito Kamata
Josh Kahn

February, 2017
Poster presented at the Pacific Coast Research Conference annual meeting

Text Messages: Examining Different Estimates of Text Complexity

Joseph F. T. Nese, Julie Alonzo, Gina Biancarosa, Akihito Kamata, Josh Kahn

Behavioral Research and Teaching, University of Oregon, 2017

Funding Source: The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A140203 to the University of Oregon.

Acknowledgements: Thank you Jeff Elmore (Metametrix) and Nia Dowell (Coh-Metrix) for your collegiality!

For more information Please contact jnese@uoregon.edu. More information on this and related projects can be obtained at *http://brt.uoregon.edu*

Purpose

Text complexity has received increased attention due to the CCSS[1], which call for students to comprehend increasingly more complex texts as they progress through grades. Quantitative text complexity (or readability) indices are based on text attributes (e.g., sentence lengths, and lexical, syntactic, & semantic difficulty), quantified by mathematical formulas to estimate text.

Researchers have warned that texts should not be created to reach particular readability scores[2, 3]; a frequent practice of "equating" ORF passages. Issues we encountered were: estimates often differ; estimates do not account for student variance; and many require texts of a certain length.

The purpose of this study was to provide comparisons of quantitative text complexity estimates based on text features to estimates of "text difficulty" based on student performance. Specifically, these comparisons are situated in the context of curriculum-based measurement (CBM) assessment of oral reading fluency (ORF), where the passages range from 20 to 105 words.

Method

Participants

910 students (Grade 2 = 259, Grade 3 = 329, Grade 4 = 322) were assessed online, via laptops in a one-to-one administration, during which each student read approximately 3 long, 5 medium, and 10 short passages.

Passages

We administered 330 ORF passages – 110 at Grades 2, 3, 4: Each passage was an original work of narrative fiction, and targeted readability at the mid-year level for each grade.

Passage Length	Words per Passage	Number of Passages per Grade
Long	≈ 85	20
Medium	≈ 50	30
Short	≈ 25	60

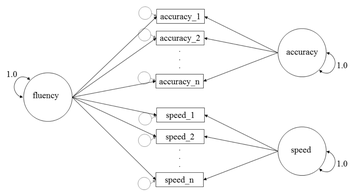
Text Complexity

Formality[4]: Many measures inform 5 principal components: narrativity, syntactic simplicity, word concreteness, referential cohesion, deep cohesion.
= (referential cohesion + deep cohesion – narrativity - syntactic simplicity - word concreteness) / 5.

Automated Readability Index (ARI): based on letter count
= 4.71(letters/words) + 0.5(words/sentences) – 21.43

Flesch-Kincaid (grade): based on syllable count (less reliable to program than letter count)
= 0.39(average sentence length) + 11.8(average syllables per word) - 15.59

WCPM (words correct per minute): Recorded audio files scored by trained assessors (with the ability to rewind, replay, and adjust audio), using the same scoring rules as traditional ORF procedures.
= (total words read – words read incorrectly) * (60 / computer recorded sec duration)



Latent fluency: Bi-factor CFA model included a general factor (fluency) and two group factors (speed and accuracy). Observed scores were at the sentence level for each reading: accuracy = percent of words read correctly, speed = average sec per word. Models estimated separately by grade.

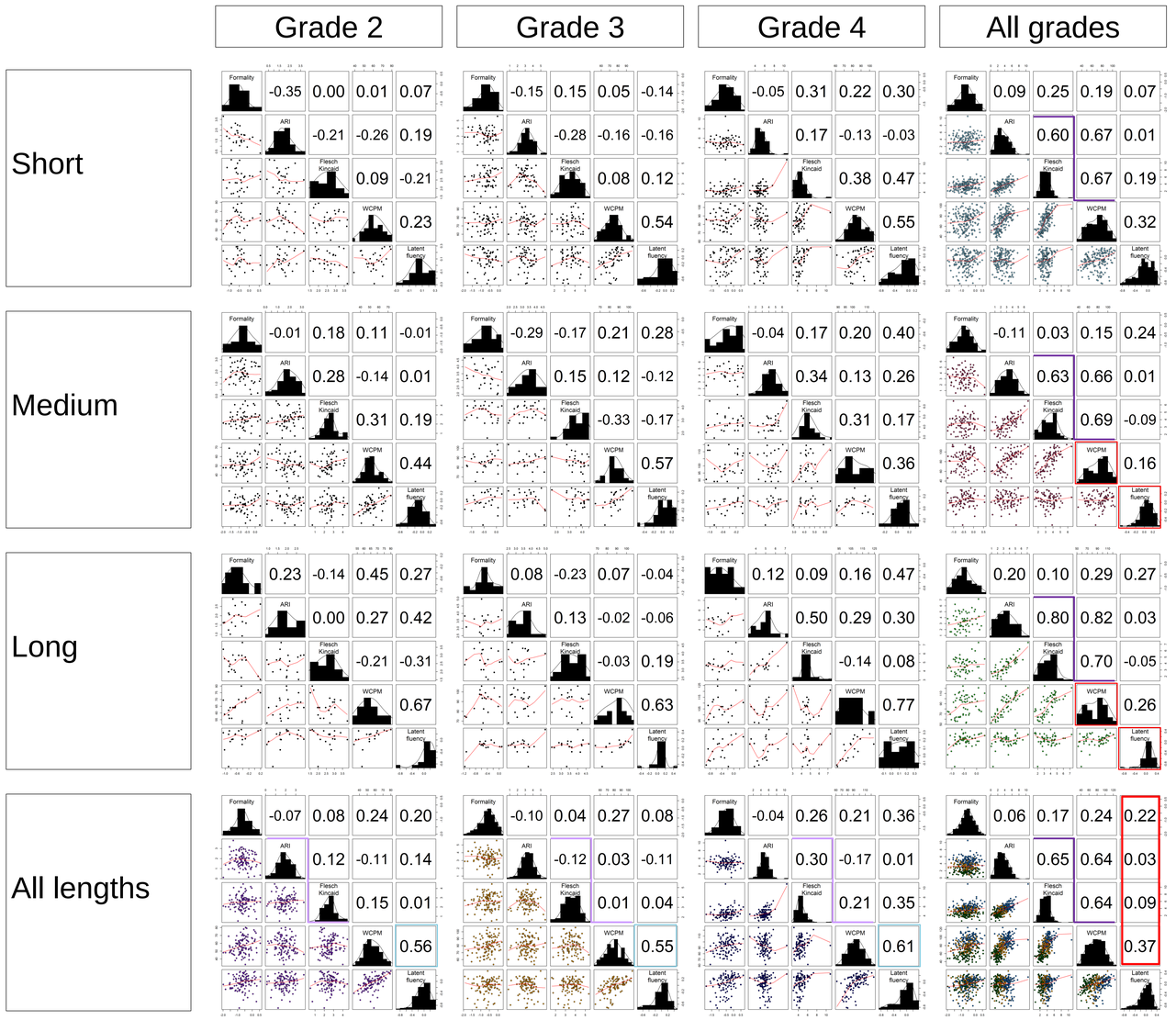
Analysis

Spearman's rank correlations were computed between the text complexity estimates. All analyses conducted with R[5, 6, 7].

Text Complexity Estimate Not Used

Coleman-Liau Index – Based on texts of 100+ words.
Gunning FOG Index – Based on texts of 100 + words.
FORCAST - Based on texts of 150 (or 100) words.
Spache – Based on text of 100-150 words & intended to measure only through Grade 3.
Dale-Chall Readability Formula – Intended to measure Grade 4 and above.
SMOG – Intended to measure Grade 4 and above.
TextEvaluator (ETS) – Ideal for texts of 300+ words.
Eight component scores as sources of comprehension difficulty (academic vocabulary, word unfamiliarity, concreteness, syntactic complexity, lexical cohesion, level of argumentation, degree of narrativity, interactive/conversational style)
Lexile – Requirement is based on text of 125+ words.
"Early Reading" Lexile – Ideal for Grade 1.
Nine text characteristics for early-grades text complexity: word structure (decoding demand and number of syllables in words), word meaning (age of acquisition, abstractness, and word rareness), and sentence and discourse-level characteristics (intersentential complexity, phrase diversity, text density/ information load, and noncompressibility).

Grade	Length	Formality		ARI		Flesch-Kincaid		WCPM		Latent fluency	
2		-0.65	(0.5)	1.87	(0.7)	2.54	(0.7)	56.88	(10.1)	-0.04	(0.2)
3		-0.62	(0.4)	3.27	(0.8)	3.58	(0.7)	79.99	(11.5)	-0.02	(0.2)
4		-0.51	(0.5)	4.90	(1.2)	4.59	(1.0)	93.67	(12.4)	-0.05	(0.3)
	Short	-0.68	(0.5)	3.40	(1.7)	3.59	(1.3)	70.67	(16.8)	-0.12	(0.3)
	Med	-0.48	(0.4)	3.22	(1.4)	3.49	(1.0)	82.55	(18.5)	0.05	(0.2)
	Long	-0.50	(0.3)	3.41	(1.4)	3.63	(1.2)	86.82	(19.2)	0.07	(0.2)
	Short	-0.75	(0.6)	1.85	(0.7)	2.54	(0.8)	52.12	(8.4)	-0.13	(0.2)
2	Med	-0.49	(0.4)	1.90	(0.8)	2.54	(0.6)	61.07	(9.6)	0.07	(0.1)
	Long	-0.58	(0.3)	1.92	(0.6)	2.53	(0.6)	64.88	(7.6)	0.07	(0.3)
	Short	-0.74	(0.5)	3.15	(0.9)	3.57	(0.8)	73.83	(9.4)	-0.08	(0.3)
3	Med	-0.47	(0.4)	3.43	(0.7)	3.60	(0.6)	85.83	(7.9)	0.02	(0.2)
	Long	-0.49	(0.3)	3.42	(0.6)	3.59	(0.6)	89.70	(10.8)	0.06	(0.2)
	Short	-0.55	(0.6)	5.20	(1.2)	4.65	(1.2)	86.07	(9.8)	-0.14	(0.3)
4	Med	-0.47	(0.3)	4.32	(1.2)	4.35	(0.7)	100.75	(7.9)	0.05	(0.1)
	Long	-0.44	(0.3)	4.88	(0.9)	4.78	(0.9)	105.88	(8.5)	0.09	(0.1)
All		-0.59	(0.5)	3.35	(1.5)	3.57	(1.2)	76.85	(19.0)	-0.04	(0.2)



Conclusions

- FK aligns closest to grade level as an artifact of the passage development
- All estimates increase across grades except: Formality remains fairly stable across Grades 3 & 4
- In our study, it was not possible to examine the Latent fluency correlations between grades because they were modeled independently. Also, WCPM and Latent fluency functioned differently by length; the former with bimodal distributions and the latter with near-normal, which was likely a function of the model choice.
- Moderate correlation among Latent fluency and WCPM, and the only consistently moderate relation within grades.
- The measure with the lowest correlations among all measures was Formality = Perhaps a different dimension of text complexity.
- Highest correlations observed among ARI, FK, and WCPM by length (which increases by length) = Partly an artifact of passage development, and partly a similarity between the formulas of ARI and FK (particularly for lower grade texts with less multisyllabic words).
- But, based on within grade correlations and previous research[8, 9, 10], this is a spurious relation resulting from the "developmental" nature of the scales.
- WCPM and Latent fluency generally increase across lengths and grades
- FK and ARI estimates remain stable across lengths

Discussion

- Does CCSS conflate text complexity with text difficulty[11]? If quantitative text complexity is a function of text features, can ORF (wcpm, latent fluency) define the interaction between text and reader ("reader-task considerations")?
- Text complexity estimates are based on entire text, but students most often do not read entire ORF passage in 60 sec.
- What is a better reflection of "**comprehension**" – measure of text or student ORF performance?

References

1. Center, N. G. A. (2011). CCSSO.(2010). *Common core state standards for English language arts & literacy in history/social studies, science, and technical subjects*.
2. Davison, A., & Kantor, R. N. (1982). On the failure of readability formulas to define readable texts: A case study from adaptations. *Reading research quarterly*, 187-209.
3. Klare, G. R. (1984). Readability. *Handbook of reading research*, *1*, 681-744.
4. Graesser, A. C., McNamara, D. S., Cai, Z., Conley, M., Li, H., & Pennebaker, J. (2014). Coh-Metrix measures text characteristics at multiple levels of language and discourse. *The Elementary School Journal*, *115*(2), 210-229.
5. R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org
6. Michalke, M. (2016). koRpus: An R Package for Text Analysis (Version 0.06-5). http://reaktanz.de/?c=hacking&s=koRpus
7. Revelle, W. (2016) psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA, CRAN.R-project.org/package=psych Version = 1.6.12.
8. Ardoin, S. P., Williams, J. C., Christ, T. J., Klubnik, C., & Wellborn, C. *(2010). Examining readability estimates' predictions of students' oral reading rate: Spache, Lexile, and Forcast. School Psychology Review*, *39*, *277–285*.
9. Compton, D. L., Appleton, A. C., & Hosp, M. K. (2004). Exploring the relationship between text-leveling systems and reading accuracy and fluency in second-grade students who are average to poor decoders. Learning Disabilities Research & Practice, ***19***, *176—184*.
10. Powell-Smith, K. A., & Bradley-Klug, K. L. (2001). Another look at the "C" in CBM: Does it really matter if curriculum-based measurement probes are curriculum-based? Psychology in the Schools, ***38***, *299–312*.
11. Cunningham, J. W., & Anne Mesmer, H. (2014). Quantitative Measurement of Text Difficulty: What's the Use?. *The Elementary School Journal*, *115*, 255-269.