

Suffixation adds complexity to children's reading, but it also serves as an effective indicator of text readability across diverse corpora.

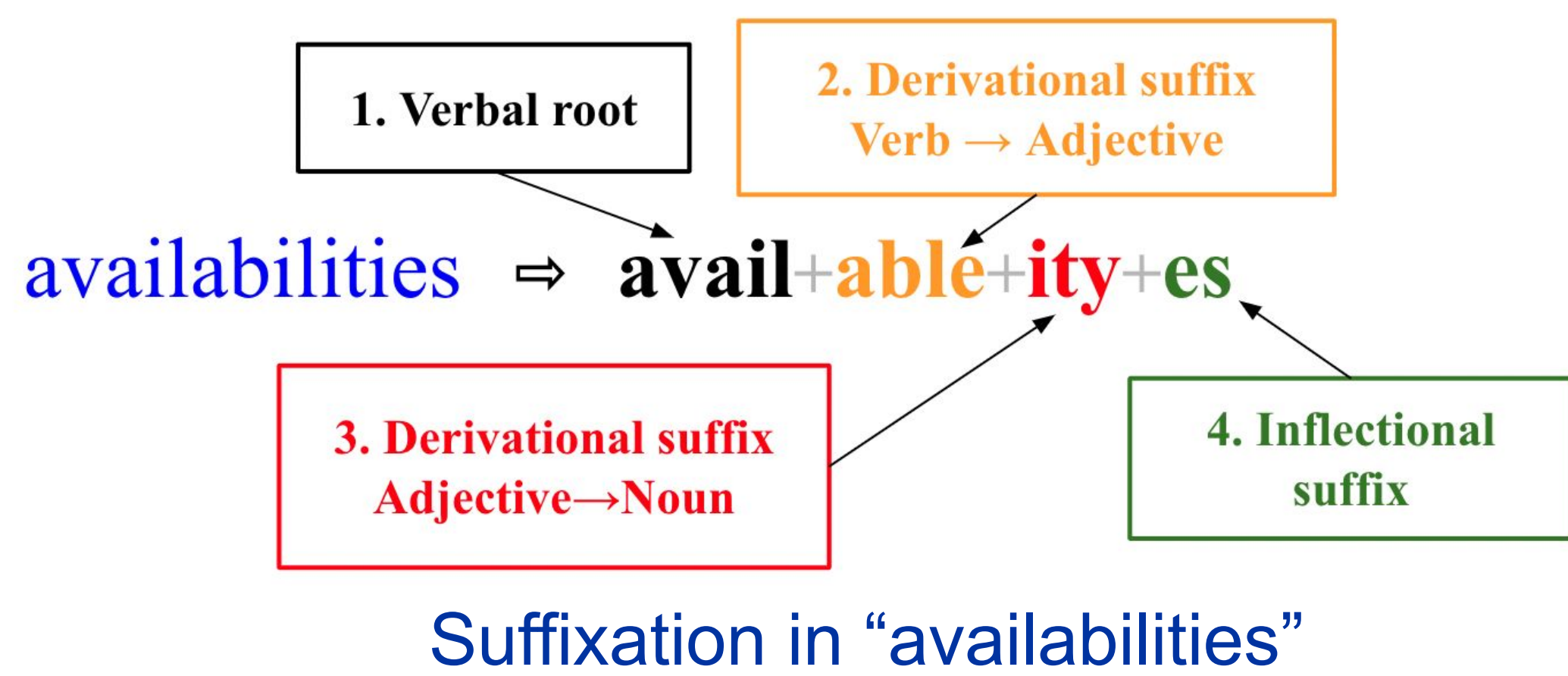
TITLE: **Beyond Vocabulary: Capturing Readability from Children's Difficulty**

AUTHOR:  **Arif Ahmed**
arifahmed@u.boisestate.edu

MOTIVATION

Children in grades K-6 use online platforms for educational and learning purposes. So, it is important to provide them with appropriate text to support effective learning. Researchers on Literacy identified "suffixation" as an influential factor that affects children's reading experience [1, 2]. So, we investigate:

Research Question: How effective are ranked suffixes from literacy research for readability assessment?

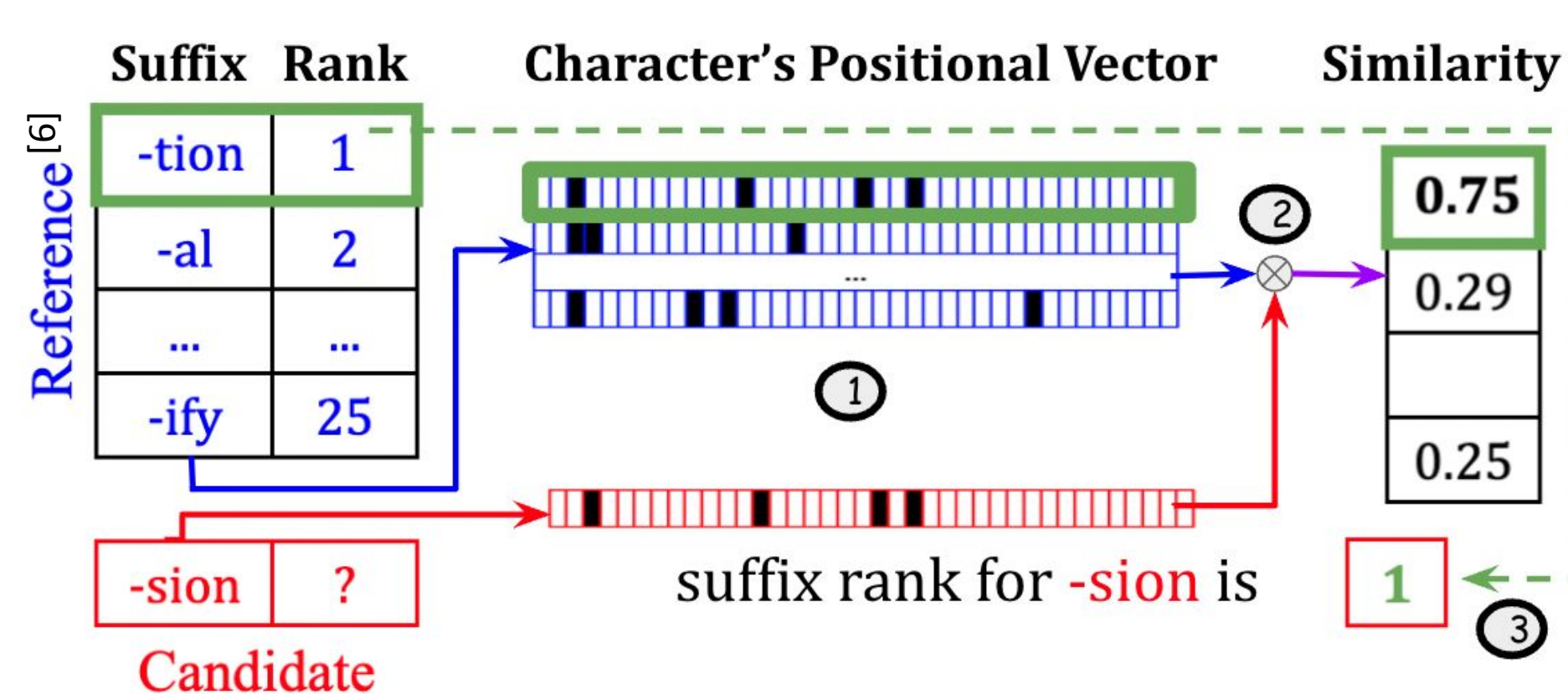


DATASETS

- Common Core State Standards (CCSS) [3]
- WeeBit [4]
- Science [5]

METHODS

Suffix Ranking for Words



Word-level Complexity

$$C_w = \begin{cases} rank & : w \text{ has derivational suffix} \\ 1 + \frac{rank}{n} & : w \text{ has inflectional suffix} \\ 0 & : w \text{ has no suffix} \end{cases} \quad (1)$$

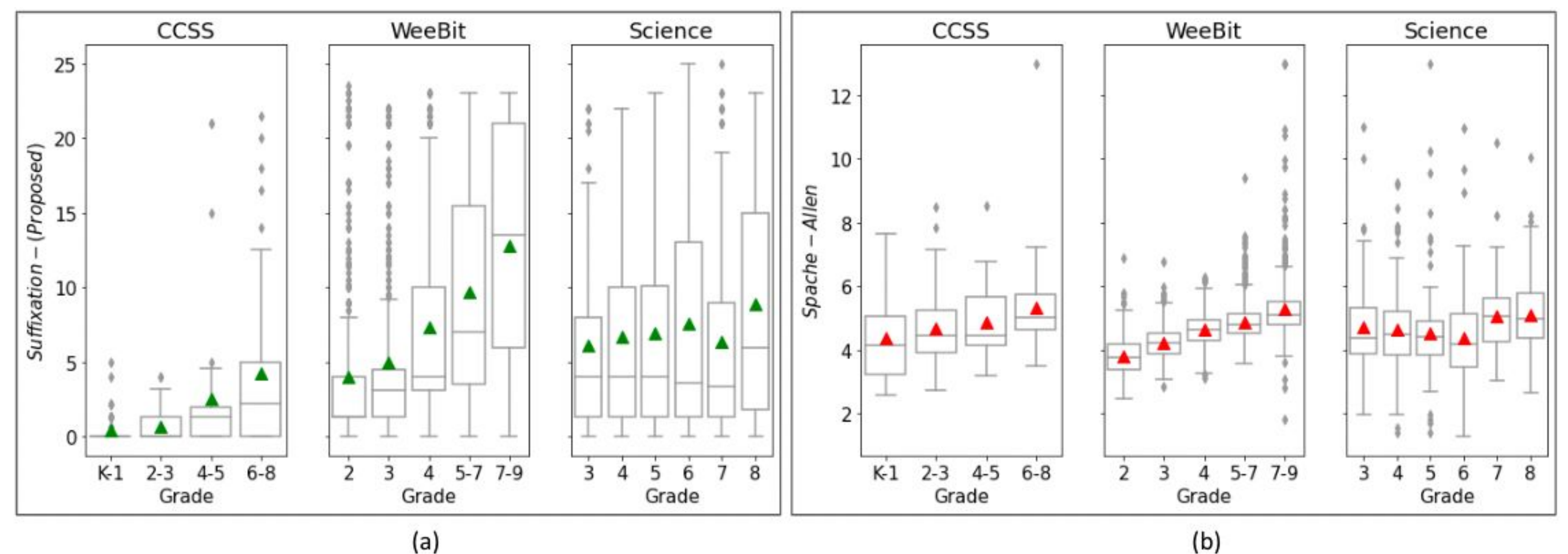
Text-level Complexity

$s1 \rightarrow$ For COVID, the ^{1.5}government ⁴aborted the ¹²municipality ⁹election.
 $s2 \rightarrow$ The new ^{1.8}chairman candidate was unhappy with this ⁹decision.

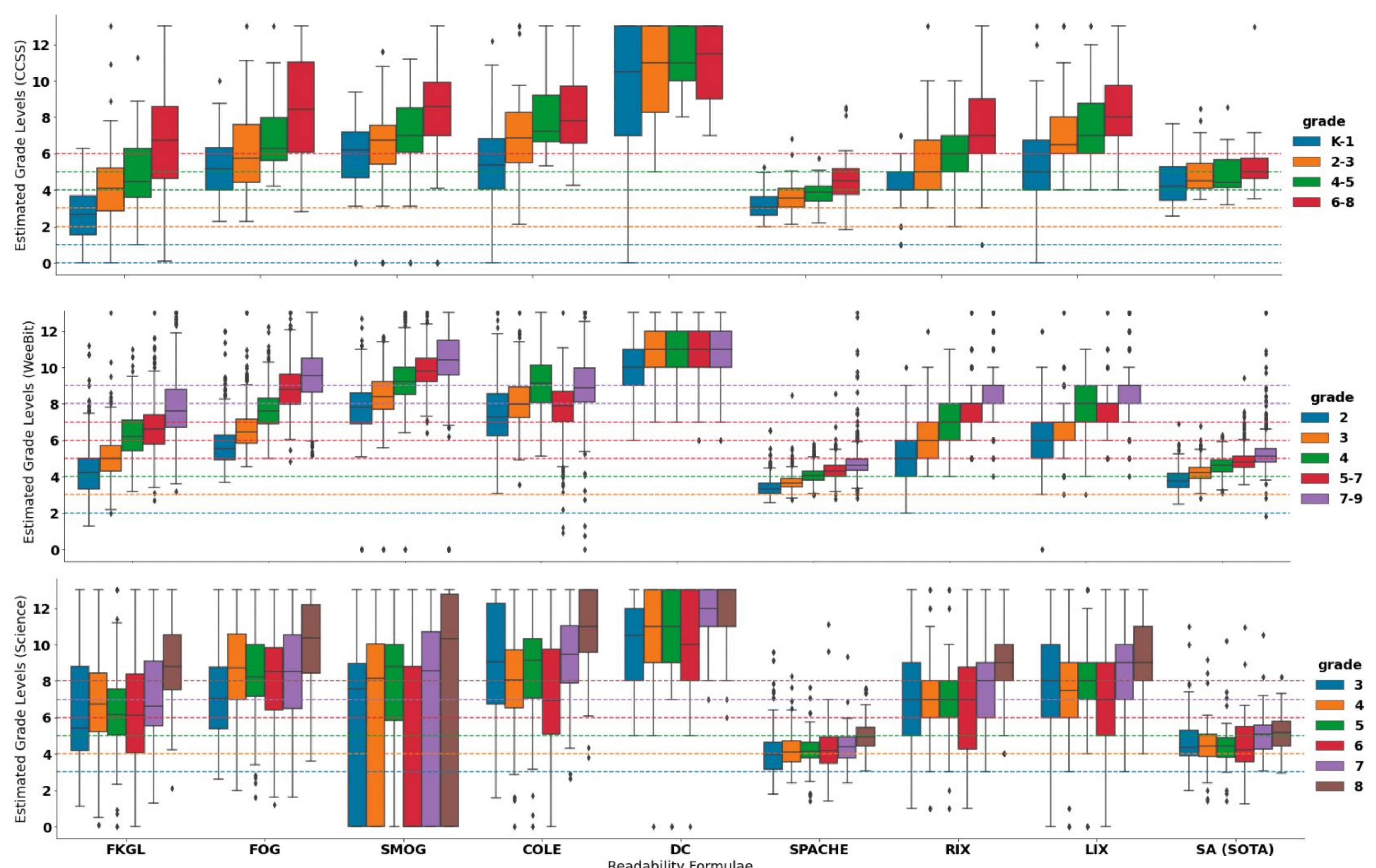
Here, we follow Equation-1 to measure word complexity. For example, $C_{municipality} = 12$

$$\begin{aligned} \therefore \text{Text-Complexity} &= \text{median} \left(\max \left(\bigcup_{w \in s1} C_w \right), \max \left(\bigcup_{w \in s2} C_w \right) \right) \\ &= \text{median} (\max([1.5, 4, 12, 9]), \max(1.8, 9)) \\ &= \text{median} ([12, 9]) = 10.5 \end{aligned}$$

RESULTS



Proposed Approach Outperforms the State-of-the-art (SOTA) Formula



Traditional & SOTA Formulae Fail to Discern Lower Grade Text

CONCLUSION

- Our novel suffixation-based approach can discern lower grade texts effectively compared to SOTA readability formula targeting children.
- The SOTA readability formula estimates higher readability scores for lower grades, such as K-2, and for texts containing simple words (e.g., cat, bat)
- Our word-level complexity scoring can directly support lexical simplification tasks, while text-level complexity scoring can enhance text accessibility for diverse user groups
- Our suffixation approach can serve as a versatile feature for feature-based models across various NLP tasks

REFERENCES

1. Nagy, W.E., Herman, P.A. and Anderson, R.C., 1985. Learning words from context. Reading research quarterly, pp.233-253.
2. Nagy, W.E., Diakidoy, I.A.N. and Anderson, R.C., 1991. The development of knowledge of derivational suffixes. CSRTF: 536.
3. <http://www.corestandards.org/assets/Appendix>
4. Vajjala, S. and Meurers, D., 2012. On improving the accuracy of readability classification using insights from second language acquisition. In Proceedings of the seventh workshop on building educational applications using NLP (pp. 163-173).
5. Nadeem, F. and Ostendorf, M., 2018. Estimating linguistic complexity for science texts. In Proceedings of the thirteenth workshop on innovative use of NLP for building educational applications (pp. 45-55).
6. Jarmulowicz, L.D., 2002. English derivational suffix frequency and children's stress judgments. Brain and Language, 81(1-3), pp.192-204.

