

# Leesbaarheid: An Analysis of Dutch Traditional Readability Formulas on Texts for Children

Joris Voogt j.voogt-1@student.tudelft.nl

Supervisor: M.S. Pera



## 1. Background

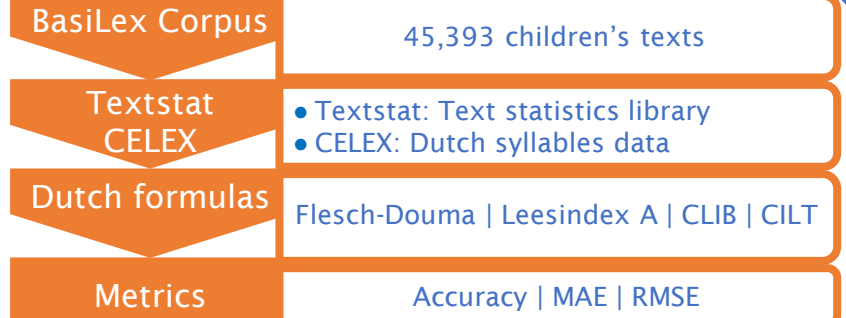
- **Children** gain knowledge through **books** and **online resources**.
- **Readability formulas** can help pick understandable texts for children.
- **Not much is known** about how **Dutch formulas** fare on texts aimed at children.
- Readability formulas were created using **physical texts**.

## 2. Research Question

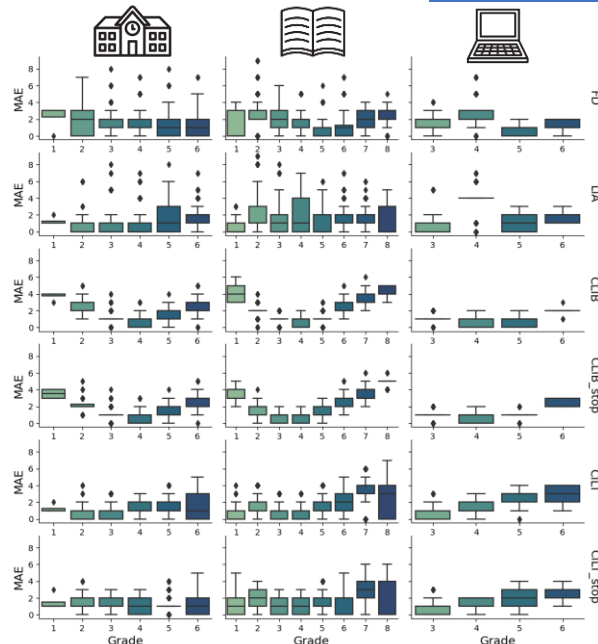
How do **Dutch traditional readability formulas** fare when estimating the readability of Dutch texts targeting **children** and is there a **significant difference** between estimations of different **types of texts**?



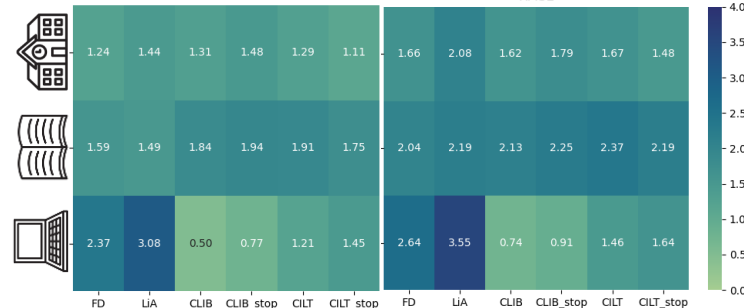
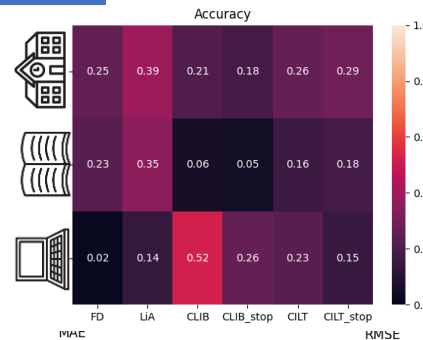
## 3. Method



## 4. Results



MAE across readability formulas, text types and grades



Metrics across readability formulas and text types

## 5. Findings

- **No overall best readability formula.**
- **CILT** best fit on school materials and books for grades 1-6.
- **CLIB** more appropriate on media, likely due to data distribution.
- **Overall significant difference** between text types.

## 6. Limitations

- Limited media entries in BasiLex
- Syllable counter based on older data set
- "Incomplete" CLIB formula
- Dated score-to-grade mappings

### References:

- "school", "book" and "laptop" pictograms by Selmon Design. Retrieved 21 May, 2023, from <https://iconicons.com/>.
- Philip van Oosten, Dries Tanghe, and Véronique Hoste. Towards an improved methodology for automated readability prediction. In *LREC*, 2010. Language Resources Association (ELRA).
- Agnes Tellings, Micha Hulsbosch, Anne Vermeer, and Antal van den Bosch. Basilex: an 11.5 million words corpus of dutch texts written for children. *Computational Linguistics in the Netherlands Journal*, 4:191–208, 2014.
- Allen et al., Supercalifragilisticexpialidocious: Why using the "right" readability formula in children's web search matters. In *ECIR*, 2022. Springer International Publishing.
- Textstat (0.7.3) <https://github.com/textstat/textstat>
- CELEX data set. (2013) <https://github.com/KBNLresearch/scansion-generator>.