



Detecting Learning in Noisy Data: The Case of Oral Reading Fluency

Beata Beigman Klebanov, ETS

Anastassia Loukina, ETS

J. R. Lockwood, ETS

Van Lyceralde, UNC Chapel Hill

John Sabatini, U of Memphis

Nitin Madnani, ETS

Binod Gyawali, ETS

Zuowei Wang, U of Memphis

Jennifer Lentini, ETS

Motivation

- In institutional educational contexts, learning of a skill is usually tracked through a series of specially designed tests.
- This is the case for tracking development in Oral Reading Fluency (**ORF**).
 - DIBELS: read 1-3 test passages aloud to the teacher
 - Administered in fall, winter, spring every year
- Can we detect improvement in ORF when children are reading a book aloud for learning and enjoyment – not reading test passages for a test?

Our proposal

- Children read a full-length book in a “relay” format – taking turns reading out loud consecutive passages from the book with a pre-recorded skilled adult narrator.
- The activity is taking place regularly as part of the curriculum, e.g., Drop Everything And Read slot.
- Children’s oral reading is recorded and sent for further analysis.

Relay Reader app



This study

- Upper elementary students in a New Jersey school read with *Relay Reader* on personal Kindle Fire 7 devices
 - 4th graders: 17
 - 5th graders: 25
- Book: "Harry Potter and the Sorcerer's Stone" (**HP1**)
- Narrator: Jim Dale (audiobook)
- Duration: 3-4 times a week, 20 minutes at a time
 - 95% of the students finished the book within 15 weeks

Measurement of ORF

- ORF is typically measured as "words read correctly per minute" (**WCPCM**).
 - Omissions, substitutions, mispronunciations are marked as errors.
 - Insertions and long hesitations are penalized indirectly through duration.
- DIBELS: Teacher times the reading, marks the errors, computes WCPCM.
 - Speech technologies have been successfully used for scoring ORF automatically, usually under quiet, monitored conditions.

Challenges in unobtrusive ORF tracking

- No control over the properties of the passages
 - It is known that *what* you read impacts *how* you read it.
 - Passages of HP1 vary widely in textual features.
- Different readers read different passages.
 - Each reader progresses through the story at a different pace.
- Substantial technical, acoustic, and behavioral noise
 - Regular class activity – special/quiet setting not feasible
 - Children read without one-on-one supervision.

The goal of this study is to assess feasibility of tracking ORF using data collected through *Relay Reader*.

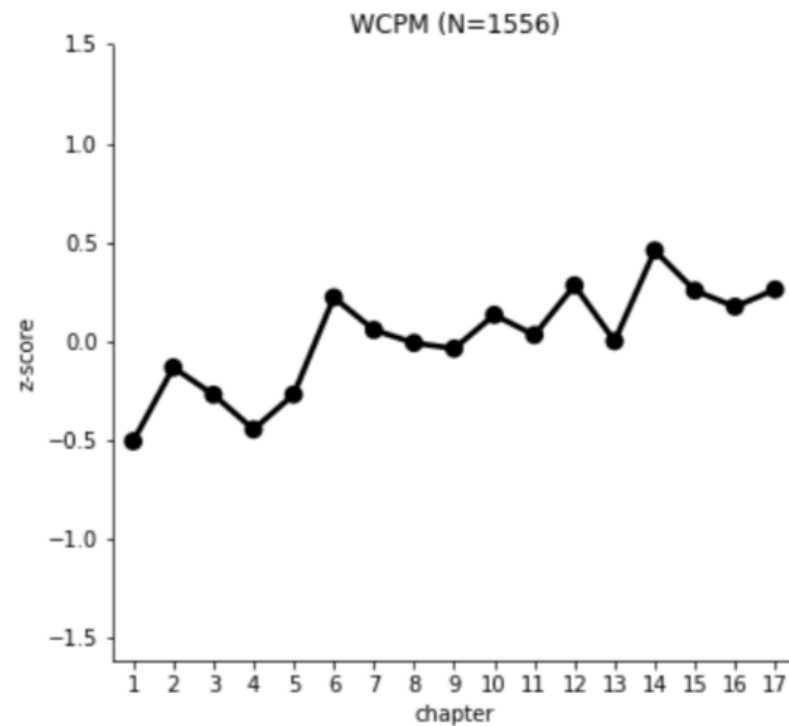
Controlling for text effects

- Text complexity estimated with Text Evaluator – **TE**
 - More complex texts are generally read more slowly
 - In ORF testing, test passages are selected or even constructed to match desired text complexity level
- Oral production in English – the words-per-minute of a state-of-art text-to-speech system - **TTS**
 - Longer words take longer to utter
 - Some phones are uttered faster than others
 - Phrase- and sentence-final lengthening and pausing

Analyses

- Treat chapters (1-17) as a measure of time.
- RQ: Is there a significant growth in ORF over time?
- Analysis 1:
 - Readers on a similar schedule (**35** out of 42 children)
 - We picked **68 passages** – 4 in every chapter – with their 1,556 recordings for **manual scoring of fluency**
- Analysis 2:
 - All data that was usable following an **automatic scoring of fluency** from **all 42 readers**

Fluency by chapter



Model

- Mixed linear model with random effects for students and for passages
- Chapter is both a fixed factor and random slope to allow students to have different ORF growth rates across chapters.

$$wcpm \sim (1|passage) + (chapter|student) + \\ + grade + TE + TTS + chapter$$

Analysis 1: Selected manually transcribed data

| <i>Dependent variable:</i> | |
|----------------------------|-------------------|
| WCPM | |
| Grade5 | -0.733 (8.909) |
| TTS | 4.841*** (0.932) |
| TE | -3.036*** (0.911) |
| Chapter | 1.278*** (0.257) |
| Constant | 99.959*** (7.510) |
| Observations | 1,556 |



Note: *p<0.1; **p<0.05; ***p<0.01

Significant chapter effect – improvement of 1.28 words correct per minute per chapter

Analysis 2: All usable automatically transcribed data

| | <i>Dependent variable:</i> | |
|--------------|----------------------------|-----------------|
| | WCPM | |
| | 35 students | 42 students |
| Grade5 | -0.40 (9.70) | 1.47 (8.71) |
| TTS | 5.60*** (0.50) | 5.78*** (0.40) |
| TE | -2.43*** (0.49) | -2.88*** (0.41) |
| Chapter | 1.21*** (0.26) | 1.12*** (0.25) |
| Constant | 62.08*** (7.91) | 60.13*** (6.77) |
| Observations | 5,182 | 6,814 |



Note: *p<0.05; **p<0.01; ***p<0.001

Significant chapter effect of the same magnitude persists

Conclusion & Future Work

- Data collected through a relay reading app can support detection of improvement in oral reading fluency using automated means, despite:
 - lack of control over text;
 - environmental and behavioral noise;
 - imperfect automated speech recognition.
- This brings us one step closer to realizing the vision of technology-based unobtrusive assessment.
- Relay Reader with more books – release 06/2021
- Encouraging early results with high schoolers in Chile



Thank you!

Please contact me at
bbeigmanklebanov@ets.org with any questions
and comments.



