Explaining Student Behaviour at Scale
The Influence of Video Complexity on Student Dwelling Time

Frans van der Sluis, Jasper H. Ginn, and Tim van der Zee
f.van.der.sluis@fgga.leidenuniv.nl
Introduction
From learning at scale
To learning about learning
Status quo

Granular analyses
- Clicks
- Seek actions
- Video length

Big predictions
- Learning outcomes
- Dropouts
- Engagement

Guo (2014): 6 minutes length optimal for engagement
Video dwelling time

A definition

– Time on task, relative to video length

Some common ideas:

– More time is a proxy of difficulty (Li, 2015)
– More time is a proxy of engagement (Guo, 2014)

Do they hold?
Research on text

Interest

- Reduces dwelling time

Difficulty

- Increases dwelling time
- But can also decrease dwelling time
Interest increases with stimulus complexity, up till the point of (in)comprehensibility (Berlyne, 1966; Silvia, 2008)
Two Hypotheses

Dwelling time is higher in videos
  – With low information complexity
  – With high information complexity

If interest indeed decreases dwelling time...
Three Challenges

1. Analyse dwelling time
2. Analyse video complexity
3. Explain dwelling time using video complexity
Challenge 1:
Analyse dwelling time
First Challenge

To derive a reliable estimator of dwelling time

– Using click-stream data
– Relative to video duration
– Based on clicks (seeks, pauses, etc.)
– Incomplete data (pauses, video ends)
Click-stream data

Our solution

- Video “sections” (in-video quizzes)
- Only students who finished watching a section
- Total of 471,179 session
- Dwelling time = Including pause-time
- Dwelling rate = excluding pause-time

<table>
<thead>
<tr>
<th>Course</th>
<th>Videos</th>
<th>Sessions</th>
</tr>
</thead>
<tbody>
<tr>
<td>configuringworld001</td>
<td>37</td>
<td>46,839</td>
</tr>
<tr>
<td>globalorder001</td>
<td>21</td>
<td>112,744</td>
</tr>
<tr>
<td>humanlanguage001</td>
<td>27</td>
<td>270,333</td>
</tr>
<tr>
<td>internationaltaxation001</td>
<td>10</td>
<td>34,213</td>
</tr>
<tr>
<td>metals001</td>
<td>6</td>
<td>6,356</td>
</tr>
</tbody>
</table>
Challenge 2: Analyse video complexity
Second Challenge

Derive a valid estimation of video complexity

- Visual and auditory information
- Account for time and speed
- Needs to explain actual behaviour!
Analyse transcripts

Our solution

- Use spoken words (feasible for MOOCs)
- Analyse video transcripts on 8 features
- Integrate multiple psycholinguistic methods
- Account for speech rate (information rate)
- Train model using Wikipedia, apply to Coursera
Features

- Traditional
- Lexical Familiarity
- Priming
- Dependency-locality
Features

F1: word length in characters
F2: word length in syllables
F3: Dale frequency
F4: Sentence length

Traditional

Feature 6: character-n-gram entropy
Feature 7: word-n-gram entropy

Priming

Feature 5: logarithm of word count per word in a representative collection (Google Books n-grams)

Lexical Familiarity

Feature 8: Integration costs of a sentence containing dependencies

Dependency-locality
Analysis of Complexity

(a) Wikipedia

(b) Coursera
Challenge 3: Explain dwelling time
Analysis of Results

dwelling rate = 3.20 − 5.37 × video complexity + 3.02 × video complexity²

$R^2 = 22.44\%$ of explained variance ($SE = 1.32, F(2,451679) = 65330, p < .001$).
Results

~22% of total variance in dwelling rate explained by complexity model

– Effects as expected
  – Low complexity makes it difficult to stay concentrated
  – High complexity troubles understanding the content
Discussion
1. Dwelling time might seem straightforward, but it’s interpretation is not.
Implications

– More variables are needed to explain dwelling time

– Dwelling rate and time are not directly interpretable and as such cannot function as a proxy measure of (perceived) difficulty nor of other related constructs.
2. Any interpretation becomes plausible once controlling for related variables.
Interpretation

Video complexity

– Strong foundation in psycholinguistics
– Difficulty affects dwelling time
– But data-driven means data-dependent
Interpretation

Interest

– Yet no direct evidence for relation to interest
3. To allow for sense-making from learning analytics, it is critical to select well-supported metrics with known theoretical relations.
Triangulation

Objective properties

Behavioral proxies <-> Subjective measures

Properties
Conclusion

Some common ideas:

– More time is a proxy of difficulty (Li, 2015)
– More time is a proxy of engagement (Guo, 2014)

Do they hold?
Thank you

For further reading: