Universiteit Leiden
The Netherlands

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

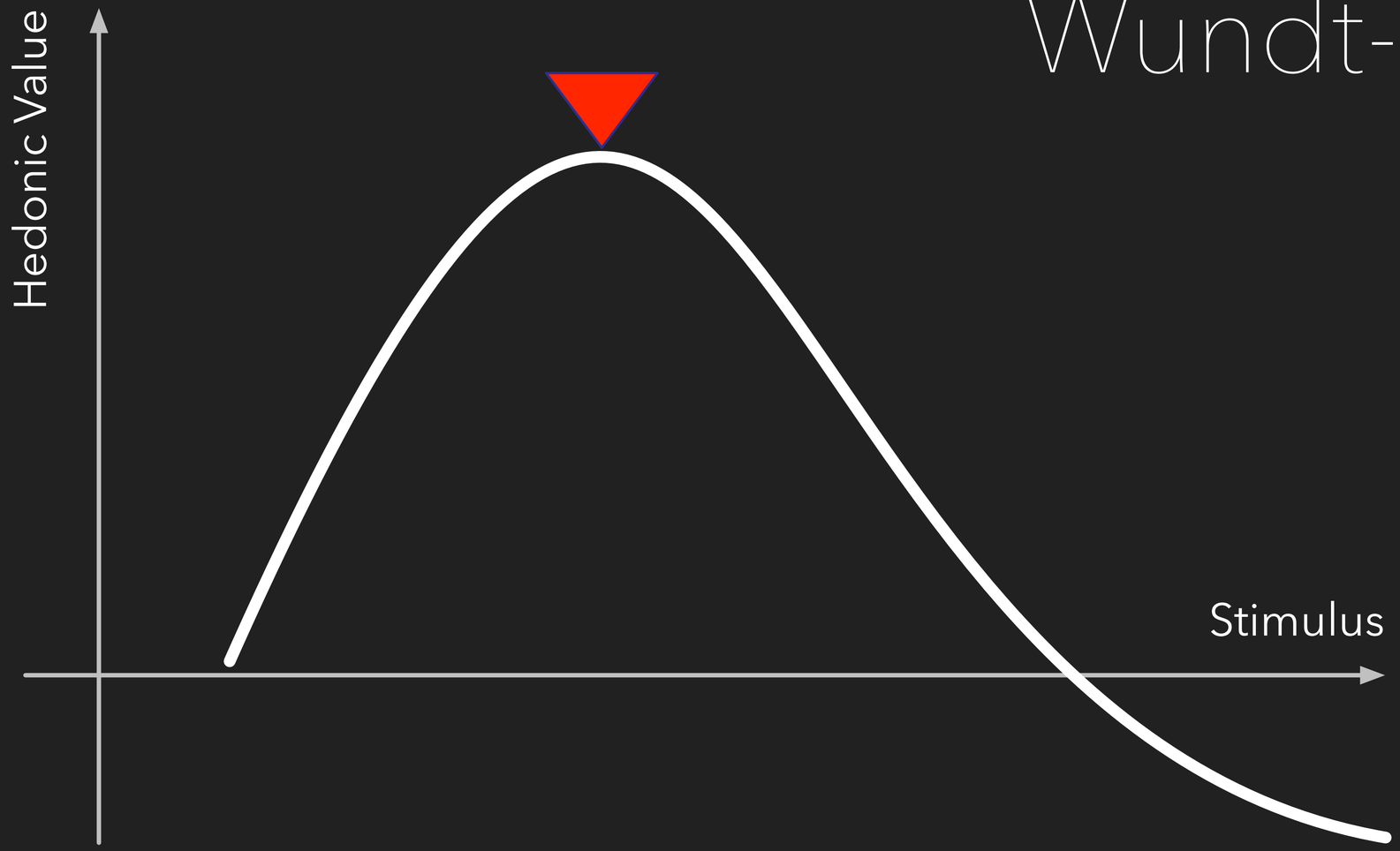
Interest

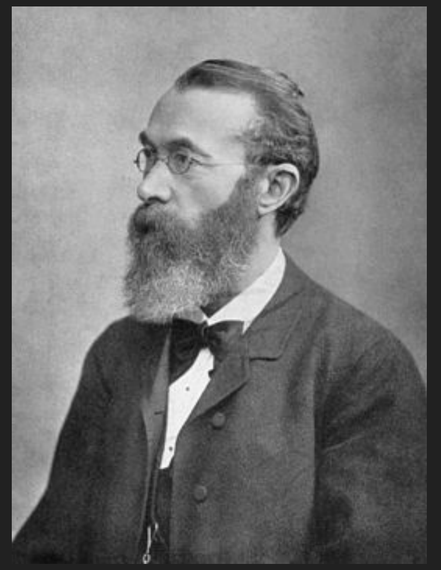
- Reduces dwelling time

Difficulty

- Increases dwelling time
- But can also decrease dwelling time

Wundt-curve (1896)





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 quizes)
- Only students who finished watching a section
- Total of 471,179 session
- Dwelling time = Including pause-time
- Dwelling rate = excluding pause-time

Course	Videos	Sessions
configuringworld001	37	46,839
globalorder001	21	112,744
humanlanguage001	27	270,333
internationaltaxation001	10	34,213
metals001	6	6,356

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

Priming

Lexical Familiarity

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

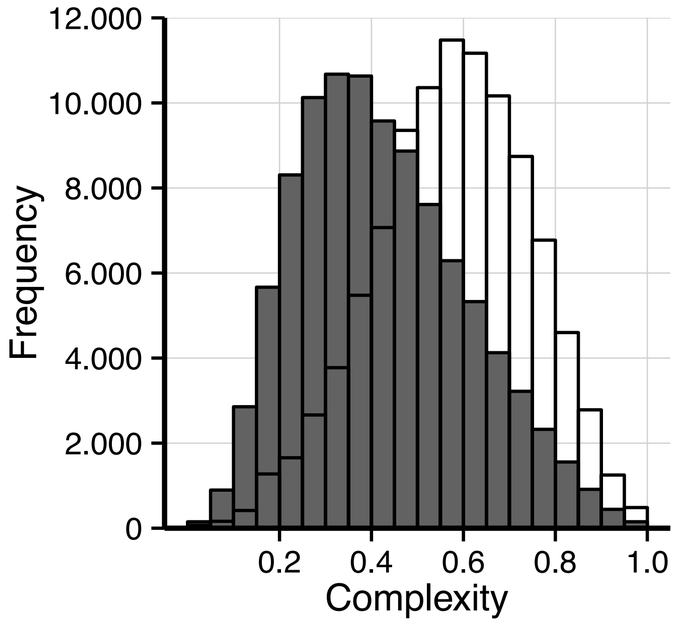
Feature 5: logarithm of word count per word in a representative collection (Google Books n-grams)

Lexical Familiarity

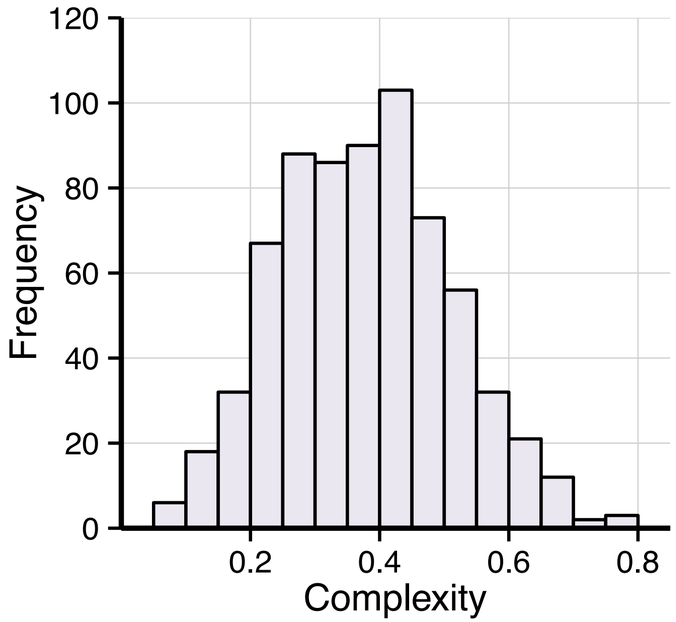
Feature 8: Integration costs of a sentence containing dependencies

Dependency-locality

Analysis of Complexity



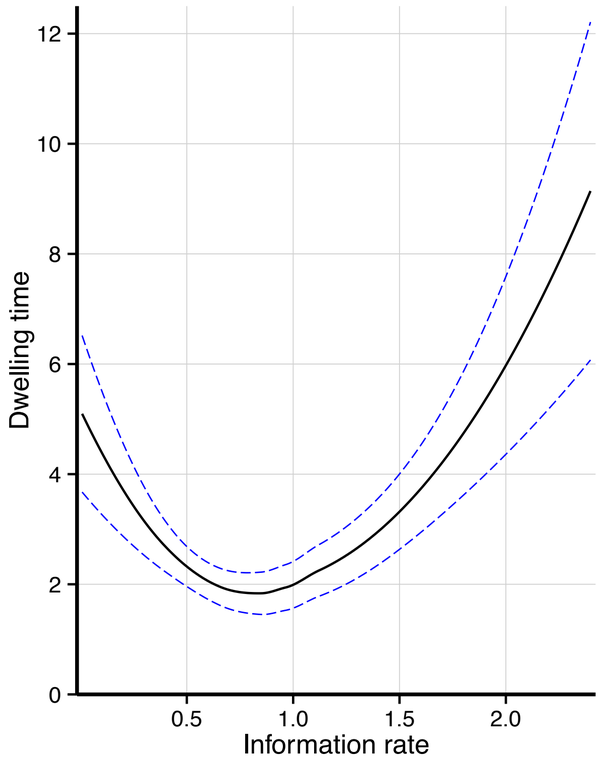
(a) Wikipedia



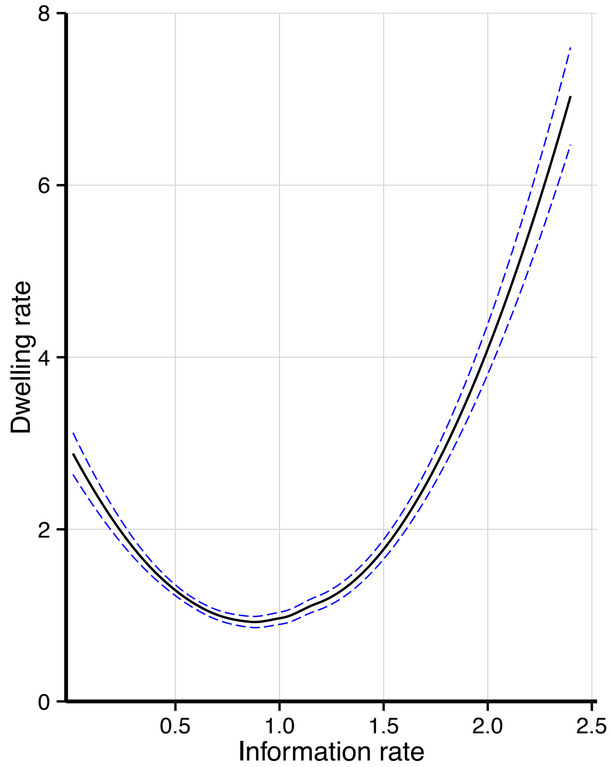
(b) Coursera

Challenge 3:
Explain dwelling time

Analysis of Results



(a) Dwelling time



(b) Dwelling rate

$$\text{dwelling rate} = 3.20 - 5.37 \times \text{video complexity} + 3.02 \times \text{video complexity}^2$$

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

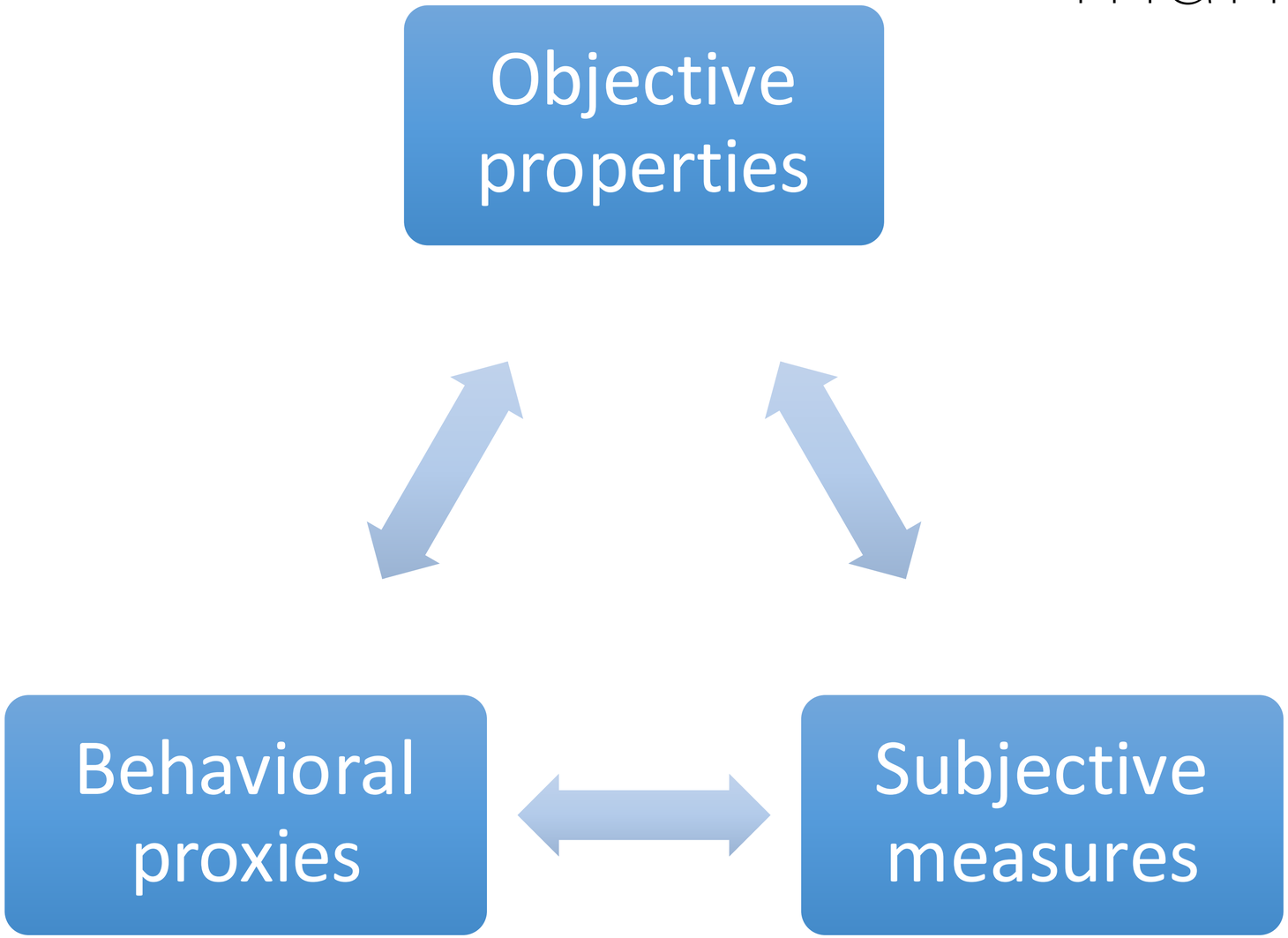
- Strong foundation in psycholinguistics
- Difficulty affects dwelling time
- But data-driven means data-dependent

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



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:

– Van der Sluis, F., Ginn, J.H., and Van der Zee, T. (2016). Explaining Student Behavior at Scale: The Influence of Video Complexity on Student Dwelling Time. In *Proceedings of the Third (2016) ACM Conference on Learning @ Scale (L@S '16)*. ACM, New York, NY, USA, p. 51-60.

– Van der Sluis, F., Van den Broek, E.L., Glassey, R.J., Van Dijk, E.M.A.G., and De Jong, F.M.G. (2014). When Complexity becomes Interesting. Journal of the American Society for Information Science and Technology 65, 7 (2014), 1478–1500.