



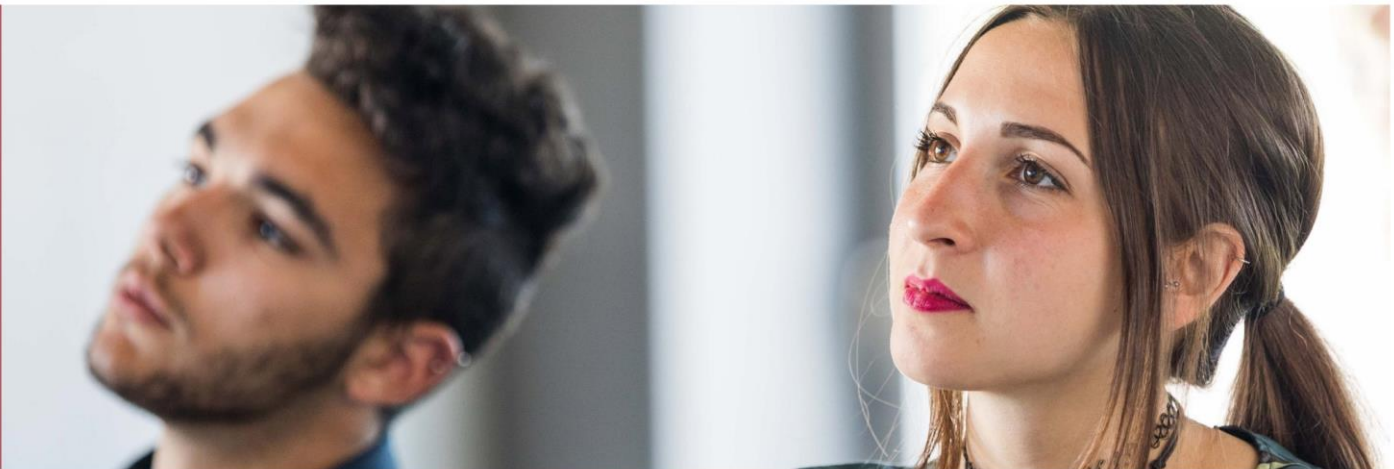
**School of Applied Linguistics**  
ILC Institute of  
Language Competence

**Curtis Gautschi**

**Predicting CEFR levels of student essays in placement tests using an Automated Essay Scoring tool in R:  
a corpus-based approach**

**THE EIGHTH INTERNATIONAL CONFERENCE ON WRITING ANALYTICS**  
**Winterthur, Switzerland: Academic Writing in Digital Contexts: Analytics, Tools, Mediality**  
[writinganalytics.zhaw.ch](http://writinganalytics.zhaw.ch)

Zurich Universities of  
Applied Sciences and Arts





# Demonstration

- Dataset
- R scripts
- Run the tool

## Background: English Writing Skills

- internationalisation of higher education
- steady increase in English-taught programmes (Wächter & Maiworm 2014)
- English as an economic (Ehrenreich 2010) and academic (Ljosland 2011) lingua franca
- writing skills: expected outcomes from students' university level education (Karras et al., 2015: 8-13)

# Placement testing

- useful to know student abilities at start of studies
- placement testing of writing tends to be avoided
- too time-consuming
- computerized text analysis: Automated Essay Scoring (AES)
- **Goal: create an AES tool to predict CEFR level of written texts**

# Setting

- ZHAW - Zürich University of Applied Sciences in Switzerland
- School of Engineering
- Online English placement test for first-year engineering students (n~600)
- Fall semester 2018



# Writing

⌚ Test time limit: 00:16:00 (ending at 10:54 AM) 00:15:11

## Instructions

🔍 Writing 0/1 📌

## Writing

You have **16 minutes** to read these instructions and write as much as you can (**minimum 100 words**).

Your text is saved automatically every minute, **but you must click "Antwort speichern" and "Test beenden" before the end of the time limit!**

**Be careful: when you click "Antwort speichern" you can't go back and change your text!!**

**Should Switzerland join the European Union? Justify your answer.**

✓ Submit answer

# AES Design

- small training corpus of texts with known CEFR levels (N=50)
- ->R->koRpus package (Michalke 2017)
- process texts (e.g., POS tagging -> treetagger)
- multiple indices (koRpus - e.g., lexical diversity, sentence length, syllables/word, word length, readability indices)
- + some of my own (e.g., ratio adj:vbs, vbs:nouns)
- prediction -> algorithm
- approach: prediction-accuracy pseudo-black box (see Vanhove et al. 2019, Yannakoudakis 2013)
- > RF regression: too good (100% accuracy) -> overfit
- > simple decision tree/boxplots/visualization -> separation at B1/B2 B2/C1
- > simple formula

# The algorithm (The Holy Grail)

CTTR x AWL x bornmuth x perc.verb / maas

Parameter	details	explanation
CTTR	Lexical diversity: Carroll's corrected Type Token Ratio	range and variety of lexical forms used
AWL	Average word length	indicator of word level/difficulty
bornmuth	Readability index	comparison of words used against common word list
perc.verb	percentage of verbs	density of verbs used in sentences
maas	Maas Lexical diversity	range and variety of grammatical structures used

$$CTTR = \frac{V}{\sqrt{2N}}$$

$$a^2 = \frac{\lg N - \lg V}{\lg N^2}$$

Where, N = tokens, V = types



Formulas Data Review View Help Tell me what you want to do

reland should join the European Union, because i can't see a big improvement for the population or the market of the count

```
71 freq.analysis(TTs[[6]], corp.freq=LCC.en)
72 for(i in 1:ncases){
```

SoE Automatic Essay Grading Sys... Grades.18HS.W Alltests.18HS.summary

	W.grade	L.score	R.score	GV.score	RLGVtots	W.level
abduli@live.com	12.337393	18	16.983	14	48.983	level 2
liv	21.111197	18	22.311	24	64.311	level 2
pas	41.118119	24	25.641	25	74.641	level 3
kev	NA	15	13.653	18	46.653	NA
a@hotmail.com	25.989354	24	23.976	19	66.976	level 3
mis	NA	24	21.978	22	67.978	NA
si1	NA	15	19.314	9	43.314	NA
med@hotmail.com	26.724270	24	NA	NA		
dmuh	NA	18	15.984	19	52	
dahm	NA	18	17.649	15	50	
tion@pascalaigner.ch	18.702356	21	23.976	26	70.976	level 2
en1	NA	18	12.654	26	56.654	NA
ard	12.551556	12	2.331	8	22.331	level 2
r	NA	18	17.649	15	50.649	NA
torfer@gmail.com	20.664588	15	15.318	9	39.318	level 2
onic	35.547464	15	24.975	21	60.975	level 3
amstutz@bluewin.ch	19.569092	24	23.976	26	73.976	level 2

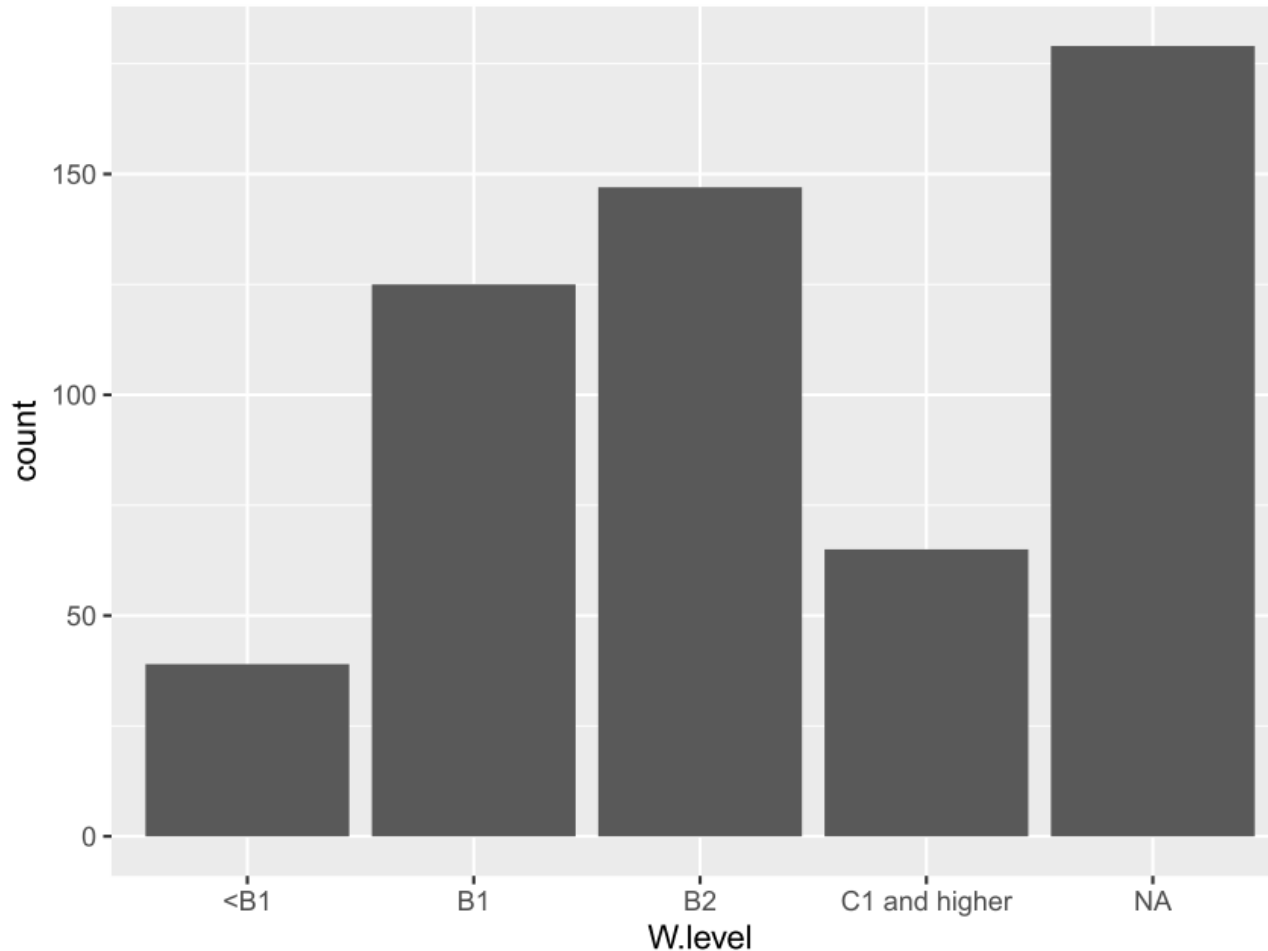
```
t(Bormuth = "/home/curtis/Tree
gger/AWL.txt", Spache = "/home
```

- Automatic output of scores in tables

# AES Results

n=370; some too short -> n=180

run-time: ~15 minutes



- **consistent with expected distributions**


# External validation: Cambridge Write&Improve vs. AES

## A review: Write and Improve

A well-known magazine for English language learners has asked for reviews from readers about useful resources for learning English.

You have recently discovered Write and Improve, a new writing tool for learners of English. Write a review for the magazine describing the main features of the tool and why it is/isn't helpful. Include details about what you like about the tool and about any difficulties you had with it, and finish by giving your recommendation for other readers.

Write your **review**.

Start again 

 Saved

A well-known magazine for English language learners has asked for reviews from readers about useful resources for learning English.

You have recently discovered Write and Improve, a new writing tool for learners of English. Write a review for the magazine describing the main features of the tool and why it is/isn't helpful. Include details about what

## Task help

 History

 Help

Level  
B1

Images

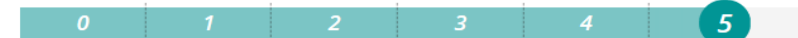
Feedback

Changes



**Great! Your writing level is B1. Now try to improve your writing to raise your level. Use the feedback to revise your writing. Then, click Check again.**

Did you write about the question? (5 is best)



A well-known magazine for English language learners has asked for reviews from readers about useful resources for learning English.

You have recently discovered Write and Improve, a new writing tool for learners of English. Write a review for the magazine describing the main features of the tool and why it is/isn't helpful. Include details about what you like about the tool and about any difficulties you had with it, and finish by giving your recommendation for other readers.

# External validation: Student texts, Cambridge texts and non-Cambridge texts (IELTS, City council citizenship tests)



## Write & Improve

		AES			
		—	B1	B2	C1
AES	B1		<b>5</b>	0	0
	B2		28	<b>19</b>	6
	C1		7	18	<b>18</b>

OA = 42% Weighted k = 0.25

Official  
Cambridge

		AES			
		—	B1	B2	C1
Official Cambridge	B1		<b>9</b>	0	0
	B2		15	<b>7</b>	3
	C1		0	5	<b>9</b>

OA = 52% Weighted k = 0.47

## Write & Improve

		AES			
		—	B1	B2	C1
Official NON-Cambridge	B1		<b>5</b>	1	0
	B2		7	<b>4</b>	1
	C1		0	2	<b>5</b>

OA = 56% Weighted k = 0.50

Official  
NON-Cambridge

		AES			
		—	B1	B2	C1
Official NON-Cambridge	B1		<b>1</b>	3	2
	B2		1	<b>7</b>	4
	C1		1	3	<b>3</b>

OA = 44% Weighted k = 0.1

# Advantages

- Written and run entirely in R
- integration with other advanced text analyses possible in R (e.g., text mining, word embedding)
- minimum of resources required
- cost-effective
- bulk grading

# Limitations

## different "genres" / text types

- For a task-independent tool: need to train on broader text types
- For a task-specific tool: need to train on that task

## design issue:

- Cambridge exams to train the algorithm (45 mins, planning)
- Test task parameters differ (write as much as you can, 15 mins -> ramble)

## black box - doesn't care about language theory

- Use the tool to explore language theory

# Applications

- Classroom-based research
  - Effectiveness of teaching methods
  - Relationships between specific text features and text quality
  - Assessment
- writing skills development -> reprogram to focus on:
  - specific skills
  - specific linguistic functions
  - specific words/work classes
  - these can be developed specifically to need rather than just using existing tools or parameters
- Integration with other advanced text analyses possible in R (e.g., text mining, word embedding)
- Natural language processing (machines)  $\leftrightarrow$  human language processing -> **controversial**

# Looking forward

## **Development:**

gold-standard (human) validation

domain/task-specific training

return to RF/decision tree regression approach

dissemination:

- shiny web app (example: <http://langtest.jp/shiny/corpus/>)
- R package





Is the cake ready?

# References

- [1] Karras, S., Konstantinidou, L., Marti, M., Müller, V., & Winkler, O. (2015). Kommunikationskompetenzen im Ingenieurberuf. Eine Bedarfsanalyse zur Förderung von kommunikativen Kompetenzen im Ingenieurstudium. Winterthur.
- [2] Ljosland, R. (2011). English as an Academic Lingua Franca: Language policies and multilingual practices in a Norwegian university. *Journal of Pragmatics*, 43, 991–1004.
- [3] Michalke, M. (2018). *koRpus: An R Package for Text Analysis*. Retrieved from <https://reaktanz.de/?c=hacking&s=koRpus>
- [4] Minsch, R., Can, E., Löhr, D., & Arqint, S. (2017). Die Fachkräftesituation bei Ingenieurinnen und Ingenieuren. *Economiesuisse DOSSIERPOLITIK*, 5, 1–18.
- [5] R Core Team. (2017). *R: A Language and Environment for Statistical Computing*. Vienna, Austria. Retrieved from <https://www.r-project.org/>
- [6] Vanhove, J., Bonvin, A., Lambelet, A., & Berthele, R. (2019). Predicting perceptions of the lexical richness of short French, German, and Portuguese texts using text-based indices. *Journal of Writing Research*, 10(3), 499–525.
- [7] Wächter, B., & Maiworm, F. (Eds.). (2014). *English-taught programmes in European higher education: The state of play in 2014*. Bonn: Lemmens Medien GmbH.
- [8] Yannakoudakis, H. (2013). *Automated assessment of English-learner writing (Technical)*. University of Cambridge, Computer Laboratory.