More than you ever wanted to know about:
Lexical diversity

Day 2

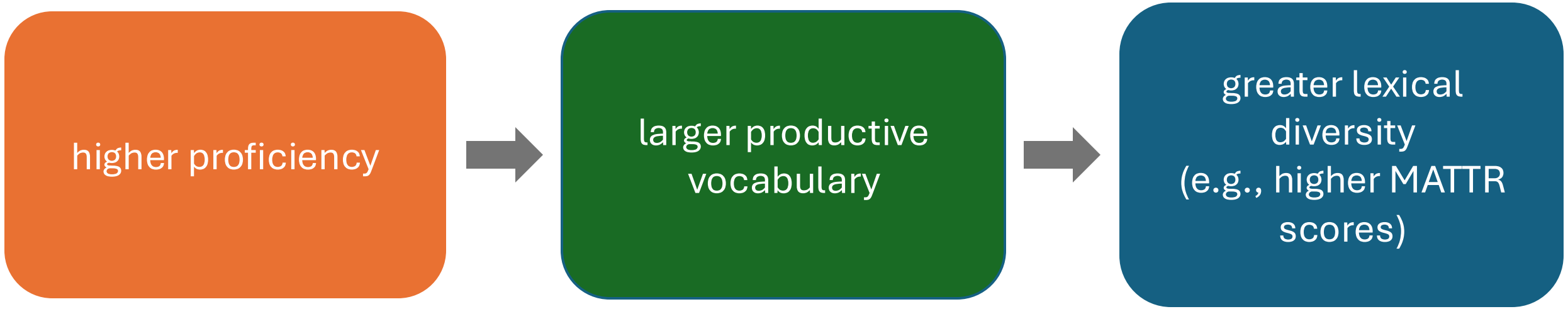
Kristopher Kyle, University of Oregon

https://github.com/LCR-ADS-Lab/

UNIVERSITY OF OREGON

Recap of Discussions So Far

- lexical diversity is one indicator of productive lexical proficiency
  - given a particular task:



Recap of Discussions So Far

A **good** index of lexical diversity will:

- be stable across different text lengths
- demonstrate a meaningful relationship with writing/speaking proficiency

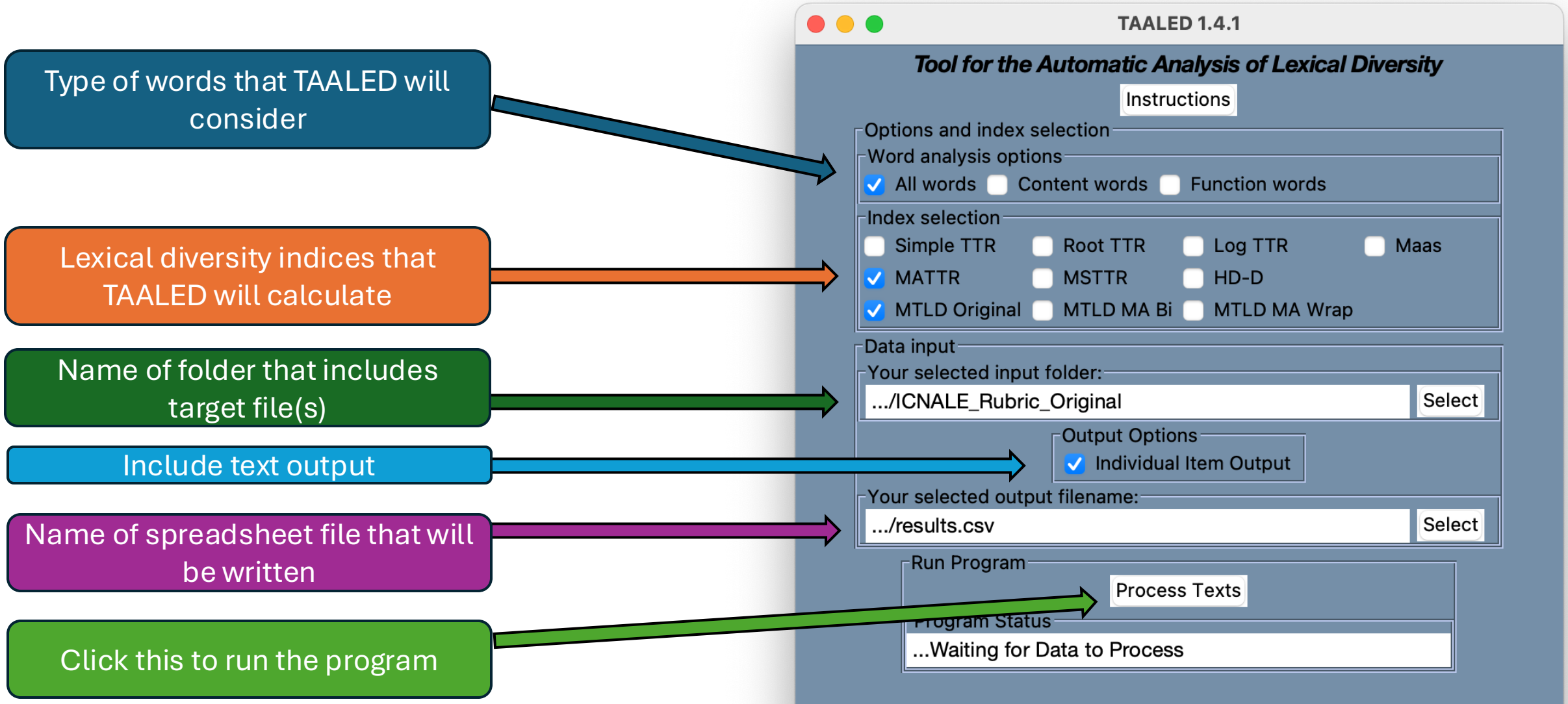
Two indices of lexical diversity that meet the criteria above:

- Moving average type-token ratio (MATTR)
- Measure of textual lexical diversity (MTLD)

Conceptual Questions?

TAALED

- Takes plain text (.txt) files (with UTF-8 encoding) as input
- Files must be placed in a folder (TAALED processes all .txt files in a folder)
- Output includes:
  - Spreadsheet with lexical diversity scores for each text
  - A processed version of each text (to check how TAALED processed each file)


Type of words that TAALED will consider
Lexical diversity indices that TAALED will calculate
Name of folder that includes target file(s)
Include text output
Name of spreadsheet file that will be written
Click this to run the program
TAALED 1.4.1
Tool for the Automatic Analysis of Lexical Diversity
Instructions
Options and index selection
Word analysis options
All words
Content words
Function words
Index selection
Simple TTR
Root TTR
Log TTR
Maas
MATTR
MSTTR
HD-D
MTLD Original
MTLD MA Bi
MTLD MA Wrap
Data input
Your selected input folder:
.../ICNALE_Rubric_Original
Select
Output Options
Individual Item Output
Your selected output filename:
.../results.csv
Select
Run Program
Process Texts
Program Status
...Waiting for Data to Process

Your text + Your lexical diversity score

Investigating the lexical characteristics of young EFL students' writing across different test tasks

Sung, Wolf, Suhan, & Kyle (2025; Assessing Writing)

UNIVERSITY OF OREGON

Overview

01	Lexical richness and L2 writing proficiency
02	Young L2 learners
03	TOEFL Junior writing test (ETS)
04	Methods, Research questions, Findings
05	Summary

Key construct: Lexical richness

- **Lexical richness** refers to the characteristics of word use (Laufer & Nation, 1995; Lu, 2012; Kyle, 2019; Yule, 1944).
  - "the number of words in a particular author's vocabulary" (Yule, 1944)
  - Lexical richness = *f*(diversity, sophistication, density) (Kyle, 2019)

Lexical richness and L2 writing proficiency

- **Lexical richness**
  - offers a framework for assessing productive lexical proficiency.
  - (as greater lexical richness is generally associated with higher writing proficiency).
  - also provides a basis for evaluating productive writing proficiency.
  - **Thus, the relationship between lexical richness and L2 writing has been widely investigated.**

Lexical richness and L2 writing proficiency

- Previous research often focused on **mature L2 learners** (e.g., Kim et al., 2018; Kyle et al., 2021; Yoon, 2018; Zenker & Kyle, 2021)

Study	Target construct	L2 learner sample
Zenker & Kyle (2021)	Lexical diversity	College students in Asia
Kyle et al. (2021)	Lexical diversity	TOEFL test takers
Kyle & Eguchi (2021)	Lexical/lexicogrammatical sophistication	TOEFL test takers
Kim et al. (2018)	Lexical sophistication	L2 learners; U.S. university
Yoon (2018)	Lexical sophistication	ESL students; U.S. university

Lexical richness and young L2 learners

- Limited focus on **young L2 learners** 

  who are broadly identified as children between the ages of 5 and 13, aligning with kindergarten through early secondary or **middle school education levels** (Wolf, 2024).

Lexical richness and young L2 learners

- Few related studies:
  - De Wilde (2023) found that lexical diversity, sophistication, and spelling accuracy predicted proficiency scores (explaining 50% of variance) among first-year Dutch-speaking secondary students.
    - diversity (MTLD ↑)
    - sophistication (e.g., bigram, trigram frequency ↑)

Lexical richness and young L2 learners

- Few related studies:
  - Wolf et al. (2018) showed that holistic TOEFL Junior Writing scores (from Korean, Portuguese learners) correlated with academic, low-frequency, and abstract vocabulary.
    - Number of academic words (↑)
    - Word unfamiliarity score (based on frequency index ↑)
    - Word concreteness score (↓)

Lexical richness and young L2 learners

- Potential challenges: **short texts**
    - De Wilde (2023): learners produced an **average of 31 words** in a picture-narration task.
    - Wolf et al. (2018): learners produced an average of **110-120 words** in an argumentative task.
        - *Standardized test setting* (TOEFL Junior writing test)

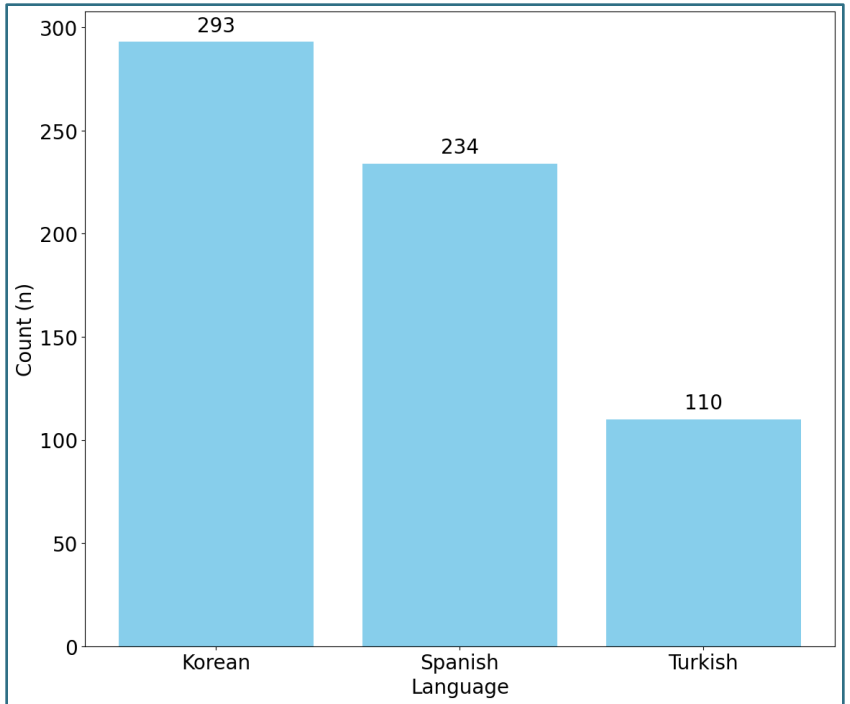
TOEFL Junior writing test

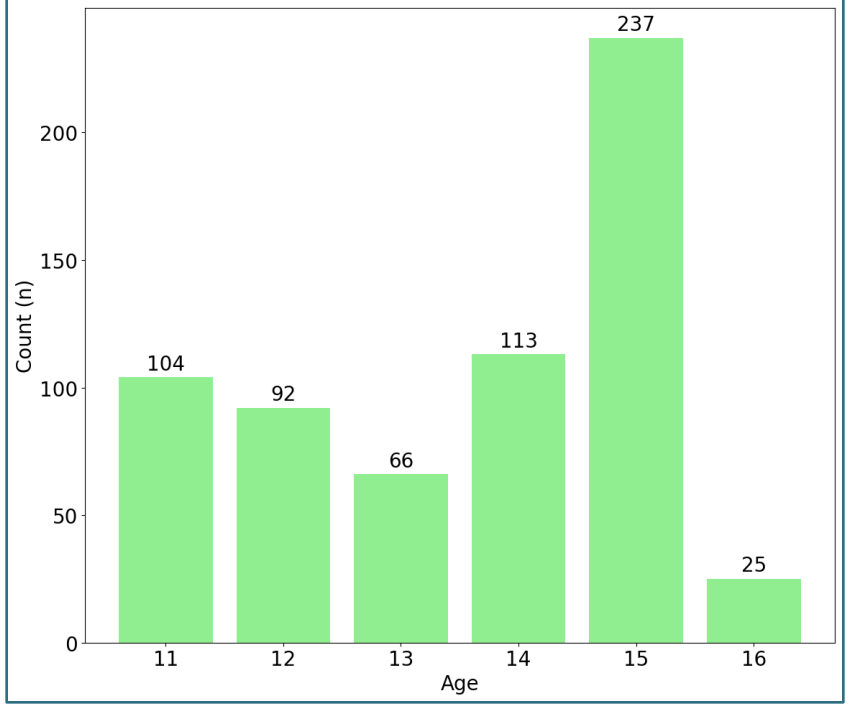
- **Construct**: computer-based English writing ability to communicate for social, school-navigational, and academic purposes
- **Target test takers**: ESL/EFL learners aged 11 and above (mostly, middle school and early high school students)
- **Four tasks**: Edit, Email, Opinion, Listen-Write

> In this task, you will hear a short talk. While you listen, you will see a picture or pictures on the screen that will help you understand the information you hear in the talk. You will be asked to write about the information that you hear.
>
> Listen to the talk carefully. You will hear the talk only one time.
>
> You may take notes while you listen. You may use your notes to help prepare your answer. You will have **10 minutes** to prepare and write your response.

Sample used in this study

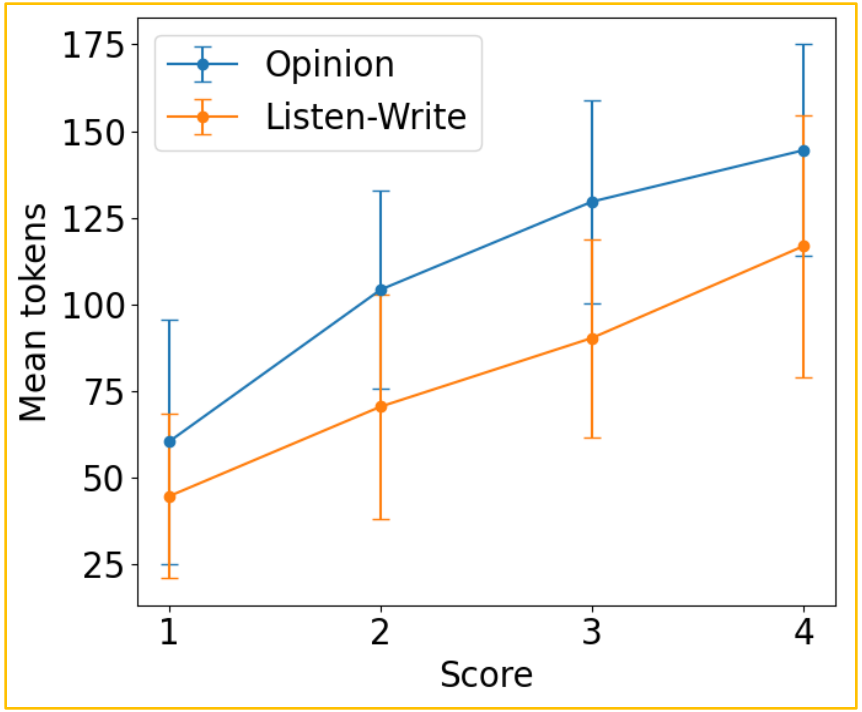
- Sample comprised 637 L2 learners from Korea, Mexico, and Turkey
- Answered the same writing prompts (Opinion, Listen-Write)
- Rated by two ETS's trained raters (from 0 to 4)





Methods

	Sub-construct	*index*
Lexical richness	**Diversity**	***optimized MATTR***



- We speculated that writing scores might be highly sensitive to the word tokens.
- Text-length stable lexical diversity index (MATTR) was optimized and used (Kyle et al., 2024; Zenker & Kyle, 2021).

Methods

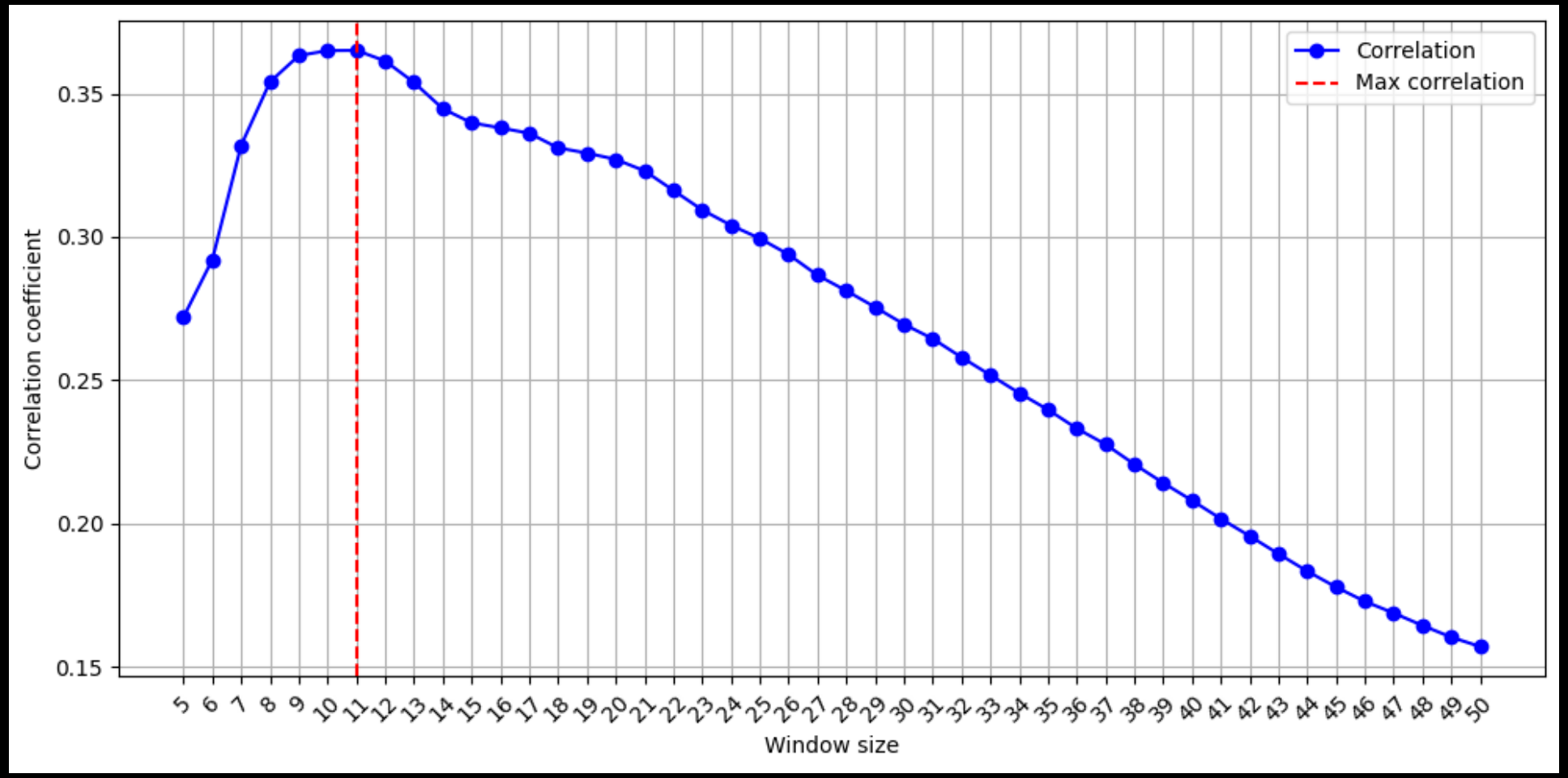
Lexical richness	Sub-construct	*index*
	Diversity	*Optimized MATTR*
	Sophistication	Multivariate approach

- *Word frequency*
- *Word range*
- *Psychолinguistic norms* (*familiarity, concreteness, imageability, meaningfulness*)
- *Age of acquisition*
- *Contextual distinctiveness*
- *n-gram use* (*frequency, range, strength of association*)

- Building on previous multivariate approaches in measuring lexical sophistication (e.g., Kyle & Crossley, 2015, 2016; Kyle et al., 2018; Yoon, 2018), we examined six dimensions of lexical sophistication.

Optimizing MATTR replicated Kyle et al. (2024)

- *Opinion writing task*



Methods

	Sub-construct	*index*
Lexical richness	Diversity	*Optimized MATTR*
	Sophistication	Multivariate approach
	Density	*the ratio of content words*

- Lexical density is often examined alongside cohesion features in L2 writing (Crossley et al., 2016).
- Some studies reported no significant relationship between lexical density and proficiency (Engber, 1995; Linnarud, 1986).

Research questions

1. What is the **relationship** between lexical richness indices and task scores in the Opinion and Listen-Write tasks?

2. How do lexical richness indices **distinguish** young L2 learners across **task scores** in the Opinion and Listen-Write tasks?

3. How do lexical richness indices **differ between task types** in young L2 learners' writing on the Opinion and Listen-Write tasks?

Some notes on Correlation

Correlations measure the relationship between two variables

Correlation values range from -1.0 to +1.0

A positive value indicates that as one variable increases, so does the other variable.

- example: temperature and ice cream sales

A negative value indicates that as one variable decreases, the other increases

- example: temperature and sweater sales

Correlation strength is indicated by the absolute value (distance from zero)

Interpreting correlation strength

Below are the guidelines from Cohen (1988):

- Small: .100 - .299
- Medium: .300 - .599
- Large: .600 - 1.000

Correlations (Opinion Task)

Table 4

Correlations of the lexical indices (Opinion).

Index	*r*	*p*
MATTR	0.385	< .001
AoA_AW	0.240	< .001
COCA_academic_tri_MI	0.210	< .001
COCA_academic_bi_MI	0.207	< .001
Brysbaert_Concreteness _AW	0.138	< .001
lexical_density_tokens	0.136	< .001
COCA_academic_Frequency_Log_AW	−0.134	< .001
COCA_academic_bi_Range_Log	−0.139	< .001
USF_CW	−0.166	< .001
MRC_familiarity_AW	−0.191	< .001

*ets

Findings (Opinion)

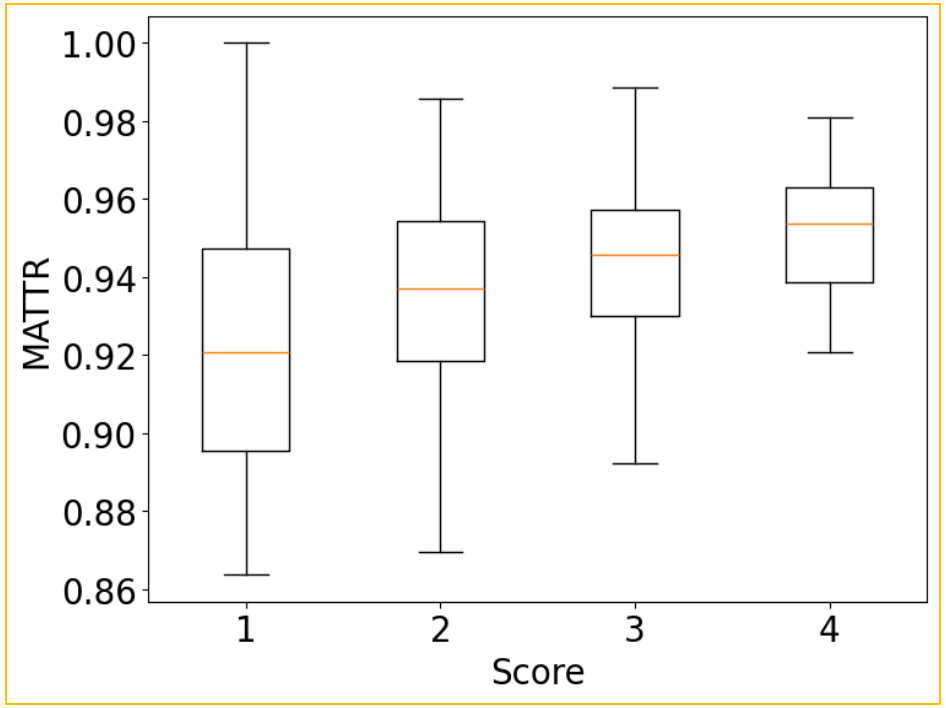
1. What is the **relationship** between lexical richness indices and task scores in the **Opinion** task?

Index	Estimates	SE	*t*	Relative importance	**Corresponding construct**
(Intercept)	-10.45	0.36	-3.72		
MATTR	7.76	0.83	9.37	0.602	**Diversity (60%)**
AoA_AW	0.56	0.13	4.49	0.167	**Sophistication (AoA) (16.7%)**
COCA_academic_tri_MI	0.08	0.04	1.94	0.086	**Sophistication (*n*-gram) (20.2%)**
COCA_academic_bi_MI	0.18	0.81	2.29	0.084	
COCA_academic_bi_range_log	-0.32	0.21	-1.50	0.032	
COCA_academic_frequency_log_aw	0.56	0.23	2.44	0.029	**Sophistication (word frequency)**

R^2 = 0.21, adjusted R^2 = 0.20

Findings (Opinion)

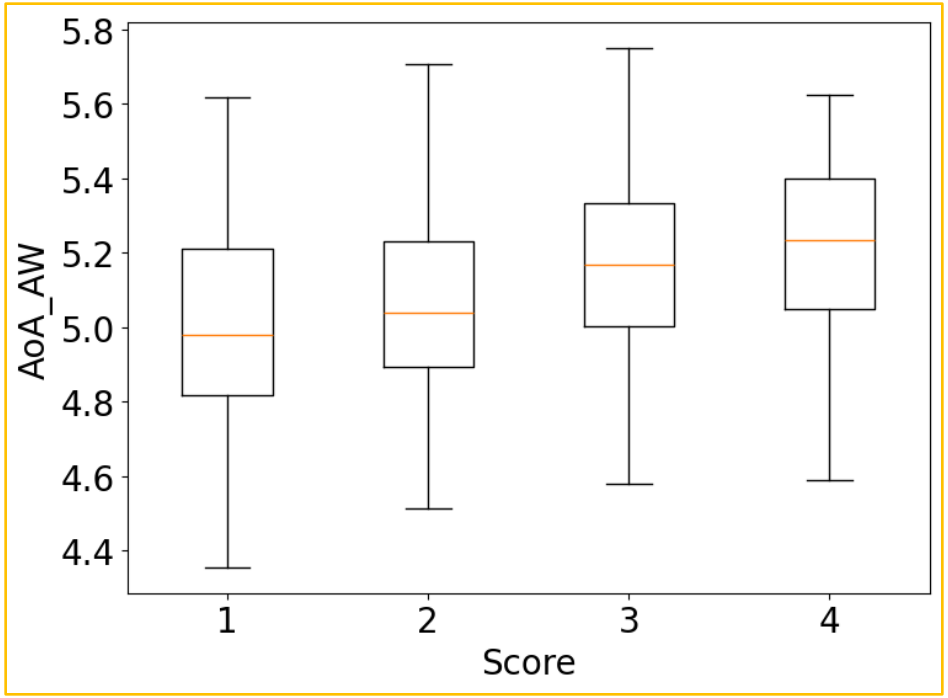
2. How do lexical richness indices **distinguish** young L2 learners **across task scores** in the **Opinion** task?



Index	F	*p*	η2	Post-hoc	Mean difference	*p* adj
MATTR	41.04	<.001	0.16	2-1	0.028	***
				3-1	0.039	***
				4-1	0.049	***
				3-2	0.011	**
				4-2	0.021	**
				4-3	0.010	

Findings (Opinion)

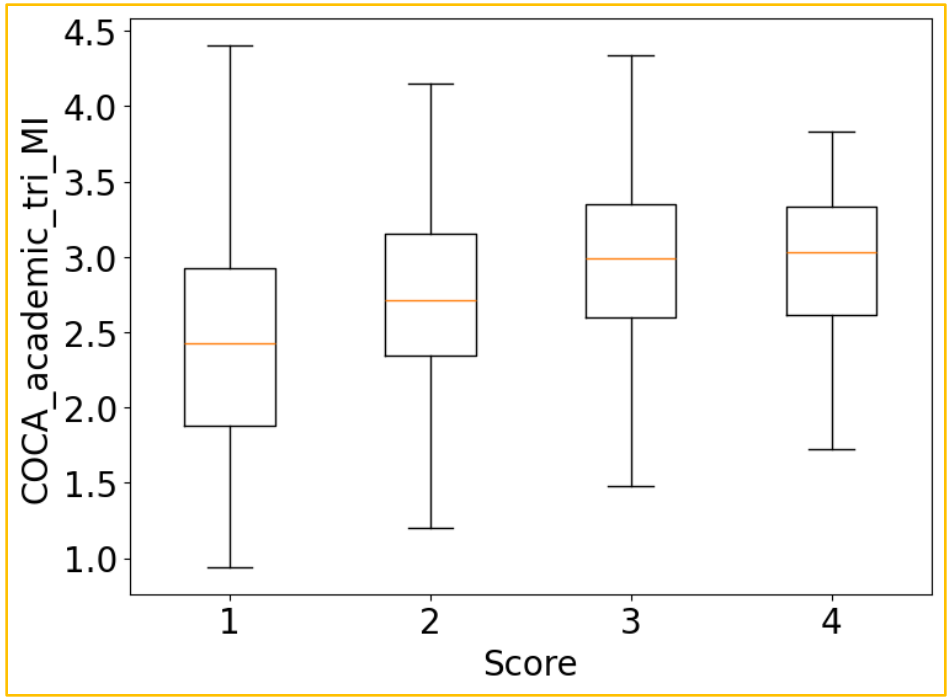
2. How do lexical richness indices **distinguish** young L2 learners **across task scores** in the **Opinion** task?



Index	F	*p*	η2	Post-hoc	Mean difference	*p* adj
AoA_AW	13.31	<.001	0.06	2-1	0.049	
				3-1	0.148	***
				4-1	0.224	***
				3-2	0.099	**
				4-2	0.175	***
				4-3	0.076	

Findings (Opinion)

2. How do lexical richness indices **distinguish** young L2 learners **across task scores** in the **Opinion** task?



Index	F	*p*	η2	Post-hoc	Mean difference	*p* adj
COCA_academic_**tri**_**MI**	11.02	<.001	0.05	2-1	0.347	***
				3-1	0.510	***
				4-1	0.566	***
				3-2	0.163	
				4-2	0.219	
				4-3	0.056	

Correlations (Listen-write)

Table 6
Correlations of the lexical indices (Listen-Write).

Index	*r*	*p*
MATTR	0.441	< .001
COCA_academic_tri_MI	0.303	< .001
COCA_academic_bi_MI	0.251	< .001
COCA_academic_tri_Range_Log	−0.144	< .001
COCA_academic_bi_Frequency_Log	−0.134	< .001
USF_CW	−0.154	< .001
MRC_Concreteness_CW	−0.160	< .001

*ets

Findings (Listen-Write)

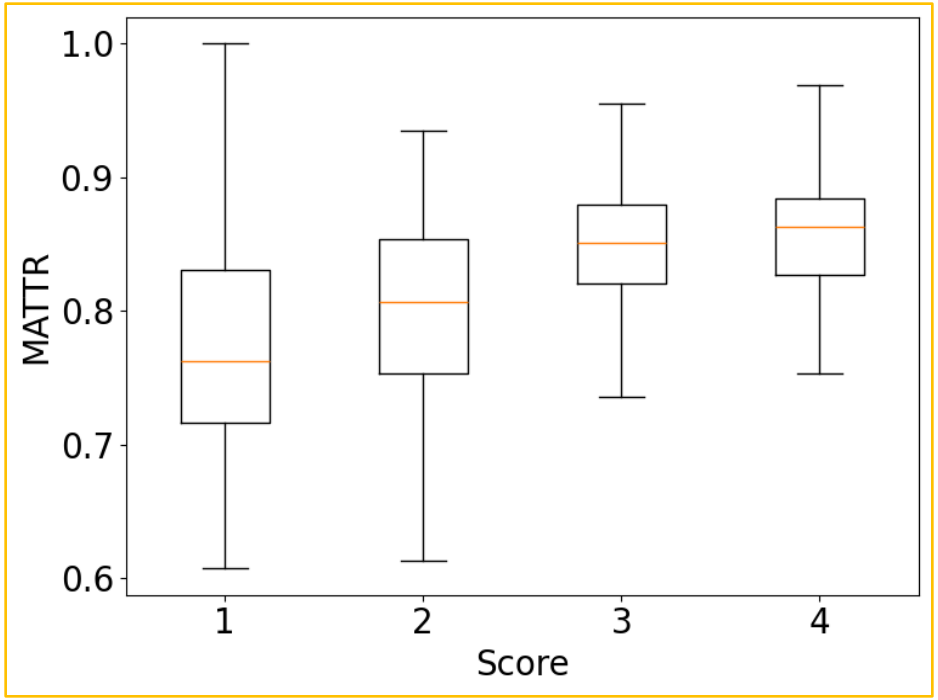
1. What is the **relationship** between lexical richness indices and task scores in the **Listen-Write** task?

Index	Estimates	SE	t	Relative importance	Corresponding construct
(Intercept)	-10.45	0.37	-3.72		
MATTR	4.42	0.43	10.24	0.622	**Diversity (62.2%)**
COCA_academic_tri_MI	0.07	0.04	1.91	0.159	**Sophistication (n-gram) (29.9%)**
COCA_academic_bi_MI	0.16	0.09	1.75	0.103	
COCA_academic_tri_range_Log	-0.11	0.07	-1.60	0.037	
USF_CW	-0.01	0.00	-3.87	0.078	**Sophistication (Contextual distinctiveness)**

R^2 = 0.25, adjusted R^2 = 0.24

Findings (Listen-Write)

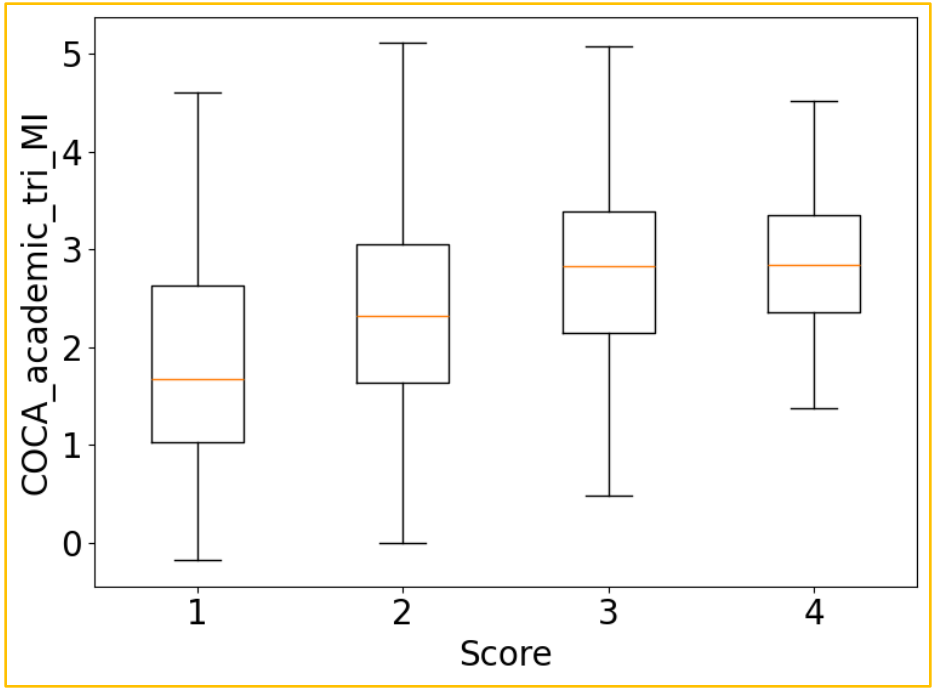
2. How do lexical richness indices **distinguish** young L2 learners **across task scores** in the **Listen-Write** task?



Index	F	*p*	η2	Post-hoc	Mean difference	p adj
MATTR	53.40	<.001	0.21	2-1	0.040	***
				3-1	0.093	***
				4-1	0.103	***
				3-2	0.053	***
				4-2	0.063	***
				4-3	0.010	

Findings (Listen-Write)

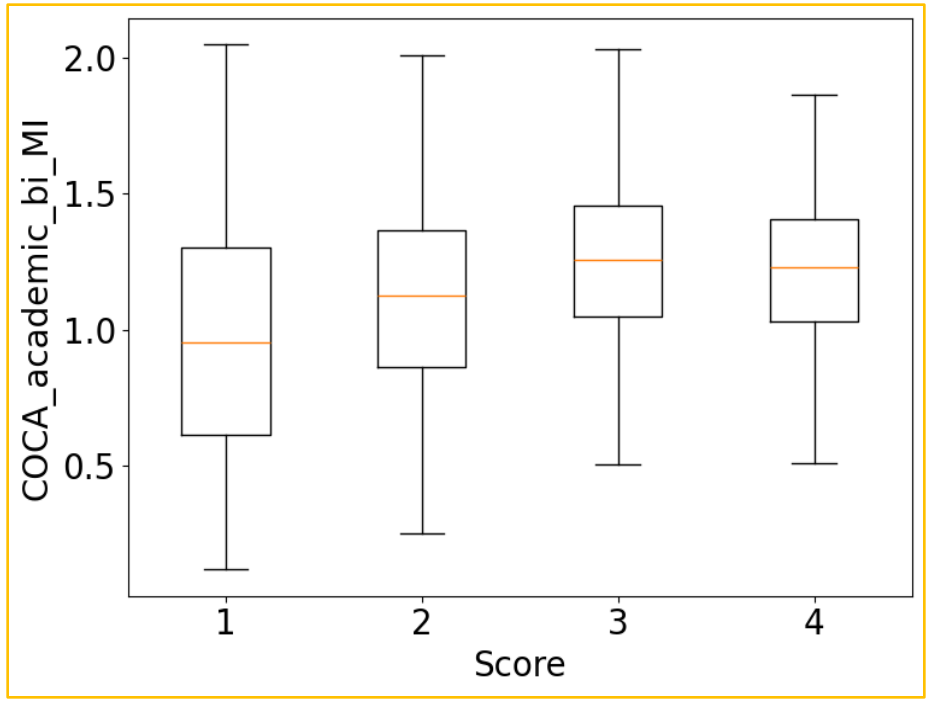
2. How do lexical richness indices **distinguish** young L2 learners **across task scores** in the **Listen-Write** task?



Index	F	*p*	η2	Post-hoc	Mean difference	p adj
COCA_academic_**tri_MI**	23.37	<.001	0.11	2-1	0.473	***
				3-1	0.015	***
				4-1	0.944	***
				3-2	0.542	***
				4-2	0.470	**
				4-3	-0.071	

Findings (Listen-Write)

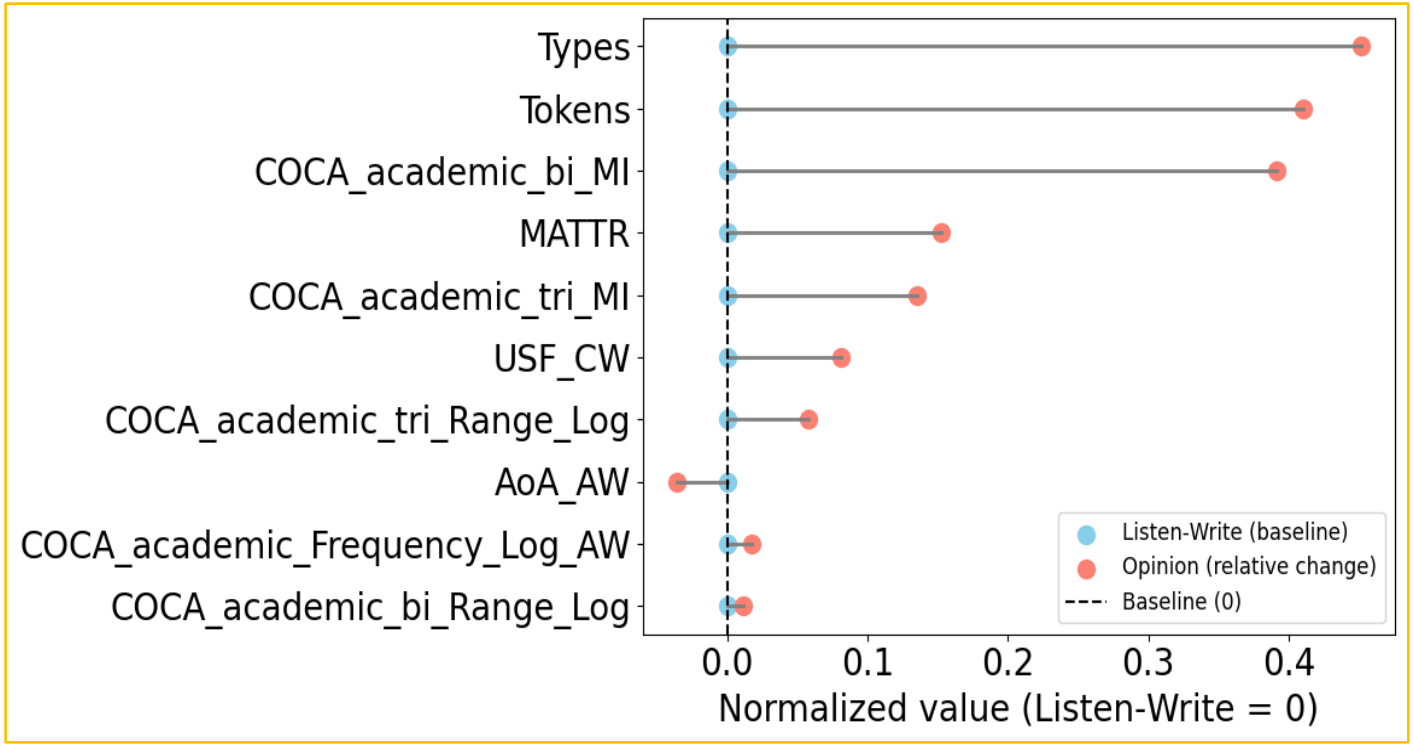
2. How do lexical richness indices **distinguish** young L2 learners **across task scores** in the **Listen-Write** task?



Index	F	*p*	η2	Post-hoc	Mean difference	p adj
COCA_academic_bi_MI	16.55	<.001	0.07	2-1	0.187	***
				3-1	0.305	***
				4-1	0.295	***
				3-2	0.118	*
				4-2	0.108	
				4-3	-0.010	

Findings

3. How do lexical richness indices **differ** between task types?



- Writings differ in *Tokens* and *Types*
- Writings from **Opinion** task overall showed higher lexical richness than the **Listen-Write** task.

Summary of findings

Lexical richness

in young L2 learners' writing

Implications

- These measures are most effective for differentiating lower/intermediate learners' writings; limited at higher scores.
  - Additional constructs (e.g., lexicogrammatical features) could be considered.

- Opinion tasks generally encourage more diverse and sophisticated vocabulary usage.
  - Listen-Write task may require more precise, source-based language; analyzing learners' vocabulary use relative to the source material may provide further insights (cf. Cai & Yan, 2024; Kyle, 2020).

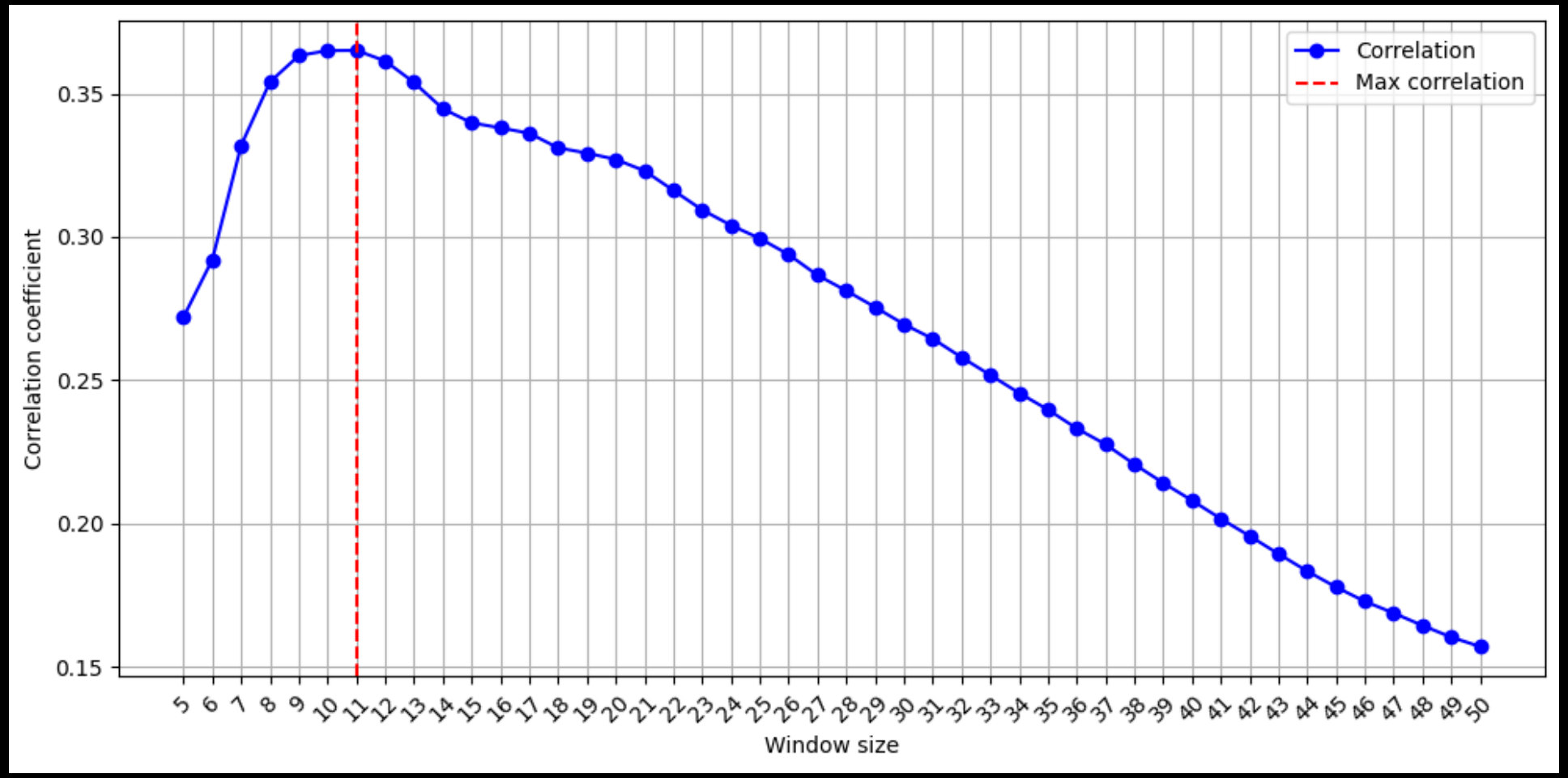
Let’s practice some correlations together!

*ets

More slides

Optimizing MATTR replicated Kyle et al. (2024)

- *Opinion writing task*



Lexical complexity – Lexical richness (Búlte et al., 2024, [illegible])

Table 1 Core and noncore measures of complexity (noncore measures in italics)

Complexity dimension/ Language level	Constitutional	Taxonomical	Hierarchical	Organizational
Lexicon	- *Mean word length*	- MATTR of lemmas - *Entropy-based measures* - *Yule's K*		

Table 2 Core and noncore measures of difficulty (noncore measures in italics)

Complexity dimension/ Language level	Causes of difficulty: Structure-related	Causes of difficulty: Context-related	Evidence for difficulty
Lexicon	- Average word length - *Derivational compositionality* - *Abstractness, imageability*	- Average frequency in reference corpus	- *Age of L1 acquisition*

ets

Findings (Opinion)

1. What is the **relationship** between lexical richness indices and task scores in the **Opinion** task?

Index	*r*	*p*
MATTR	0.385	<.001
AoA_AW	0.240	<.001
COCA_academic_tri_MI	0.210	<.001
COCA_academic_bi_MI	0.207	<.001
Brysbaert_Concreteness _AW	0.138	<.001
lexical_density_tokens	0.136	<.001
COCA_academic_frequency_Log_AW	-0.134	<.001
COCA_academic_bi_Range_Log	-0.139	<.001
USF_CW	-0.166	<.001
MRC_familiarity_AW	-0.191	<.001

Findings (Listen-Write)

1. What is the **relationship** between lexical richness indices and task scores in the **Listen-Write** task?

Index	*r*	*p*
MATTR	0.441	<.001
COCA_academic_tri_MI	0.303	<.001
COCA_academic_bi_MI	0.251	<.001
COCA_academic_tri_Range_Log	-0.144	<.001
COCA_academic_bi_Frequency_Log	-0.134	<.001
USF_CW	-0.154	<.001
MRC_Concreteness_CW	-0.160	<.001