FACULTY OF ARTS
AND PHILOSOPHY

USING CHATGPT AS A TOOL FOR AUTOMATED WRITING EVALUATION: IMPACT ON SYNTACTIC AND LEXICAL COMPLEXITY

BART DEYGERS, LIISA BUELENS, LAURA SCHILDT & MARIEKE VANBUEL

Literature review: Automated Writing Evaluation (AWE) in educational settings

Extensively studied: potential for learner autonomy and teacher workload reduction [7; 8; 24; 25]

Generative AI models (e.g., ChatGPT 3.5, 4, 4o): reliability and consistency in scoring after training [5; 19]

ChatGPT feedback may be inversely proportional to text quality [21]

Quality of human feedback may exceed ChatGPT ($p < .05$; $d = .4$-$.8$): humans more accurate, clear & more reliable for structure + content [21]

Impact on L2 gains

May lead to increased revisions and improved accuracy [4; 6; 14, 15; 16]

Real-time AWE associated with increased lexical (but not syntactic) complexity in essay revision tasks [3]

Meta analysis: medium effect ($g = 0.55$) of AES on writing performance, but intervention conditions varied [8]

Impact of AWE vs human raters on writing product: delayed post-tests have shown advantage for AWE for accuracy [16]

Challenges and Limitations

AWE constructs may differ from those valued in classroom settings; longitudinal research needed on impact in classroom settings [10; 16]

Need to examine use of generative AI on complexity, accuracy and fluency, over time [3]

Research question

Does the use of ChatGPT during a 6-week intervention in EFL writing classes lead to measurable between-goup differences regarding syntactic & lexical complexity?

Participants & design

1st year university students: translation studies, sem. 2

Course: English writing, CEFR: B2

	W1		W5	W6		W8	W9
Experiment (n = 62)		Learning path: AWE	ChatGPT	ChatGPT	ChatGPT	ChatGPT	~~ChatGPT~~
	Baseline essay	Instruction	Practice essay Targeted VLT	Essay 1	Instruction & practice	Essay 2	Essay 3
Control (n = 34)							

For each essay, factsheets were provided, to minimize role of background knowledge

No sig. between-group differences for exposure to English, age, multilingualism, self-assessed overall English proficiency, self-assessed English writing proficiency

Analysis

R (4.1.1): lme4[2], psych[19], ggplot2[21], sjPlot [18]

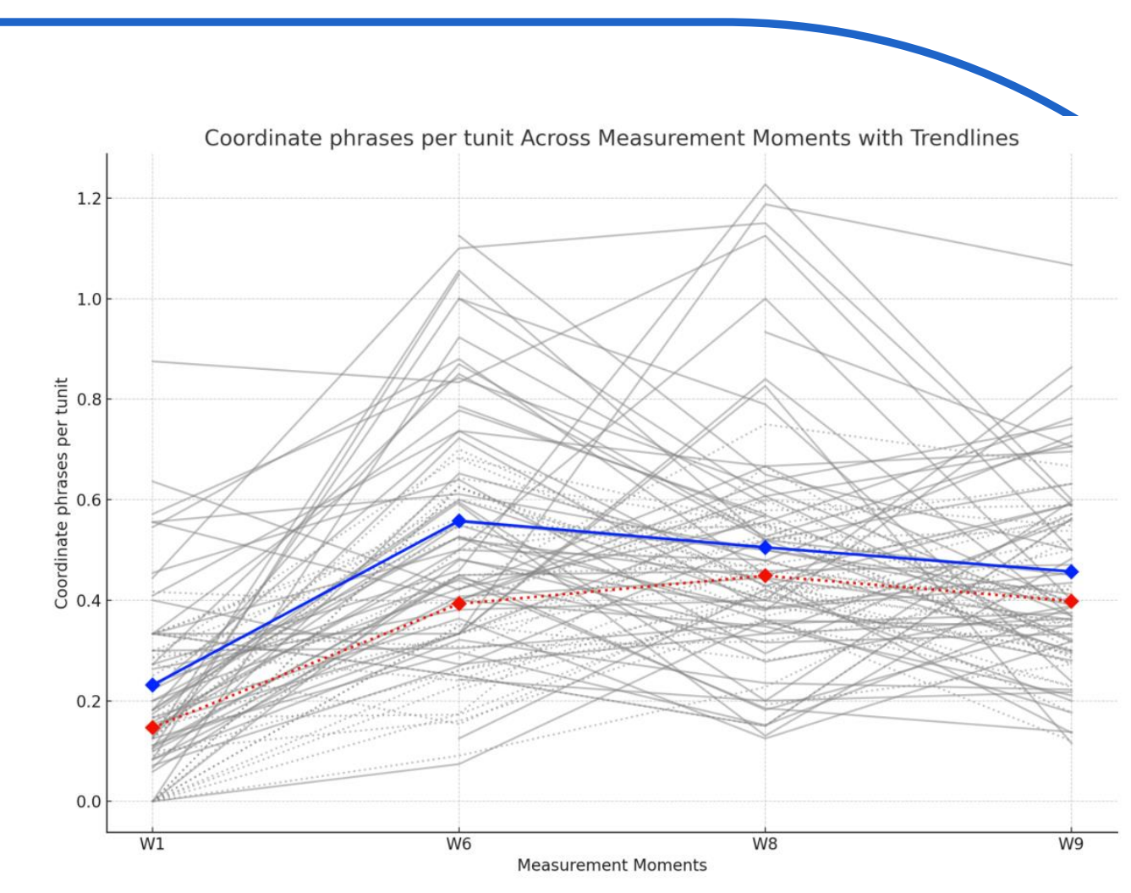
```
Model: *outcome variable * ~ 1 + week + (1 + week|ID)
```

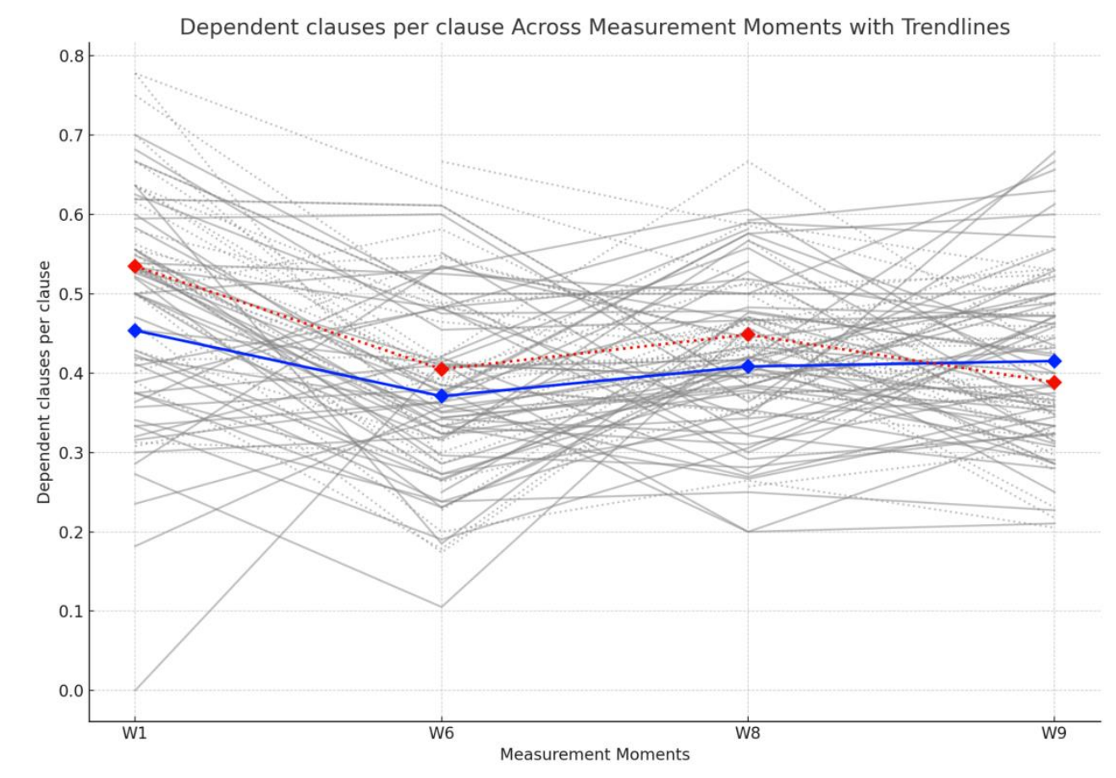
Syntactic complexity measures [11; 17]

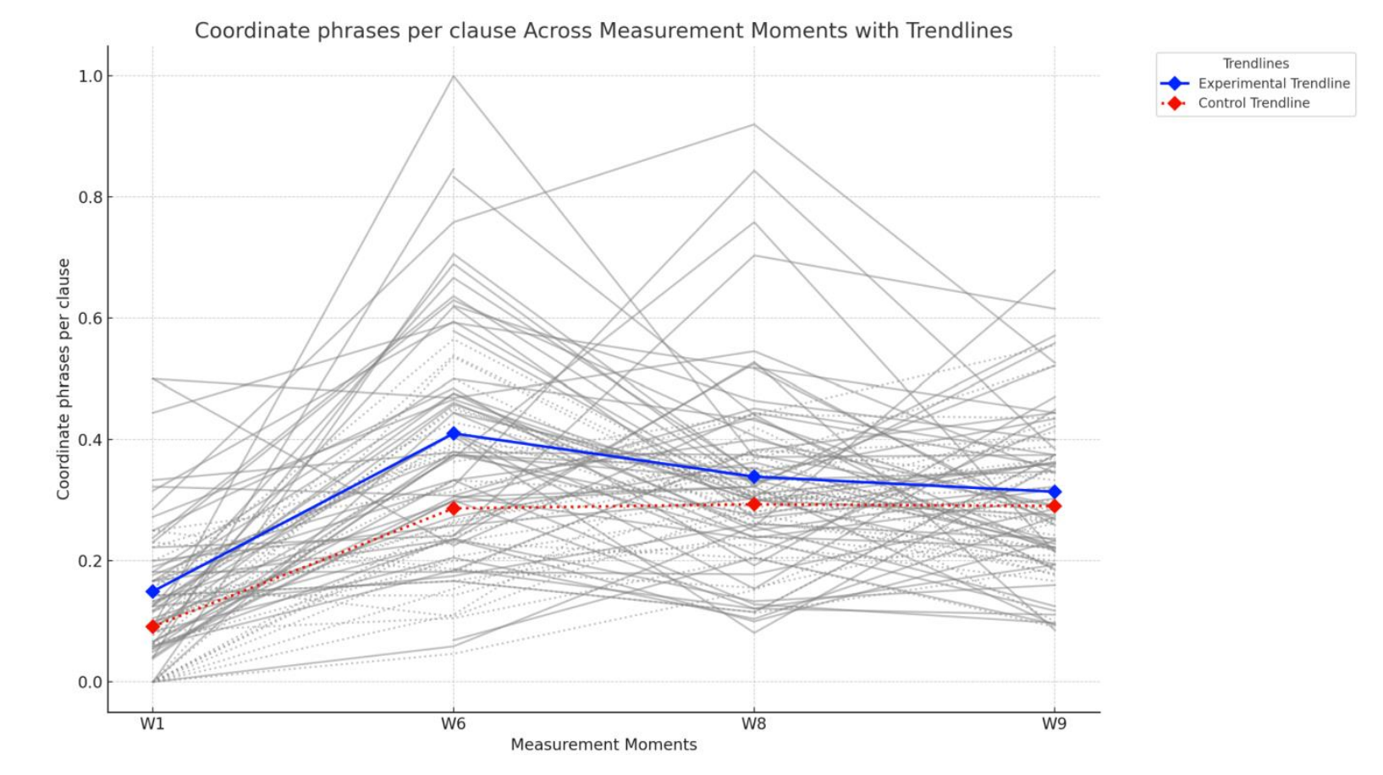
mean_length_of_clause; mean_length_of_sentence; mean_length_of_tunit; clauses_per_tunit; complex_tunit_per_tunit; dependent_clauses_per_clause; dependent_clauses_per_tunit; coordinate_phrases_per_clause; coordinate_phrases_per_tunit; verb_phrases_per_tunit; complex_nominals_per_clause; complex_nominals_per_tunit; verb_phrases_per_tunit; clauses_per_sentence

Lexical complexity measures [12; 13]

Kuperman_AoA_AW; Kuperman_AoA_CW; Kuperman_AoA_FW; Brysbaert_Concreteness_Combined_AW; Brysbaert_Concreteness_Combined_CW; Brysbaert_Concreteness_Combined_FW; SUBTLEXus_Freq_FW_Log; BNC_Written_Freq_AW_Log; BNC_Spoken_Freq_AW_Log; BNC_Written_Freq_CW_Log; BNC_Spoken_Freq_CW_Log; BNC_Written_Freq_FW_Log; BNC_Spoken_Freq_FW_Log; All_AWL_Normed; LD_Mean_Accuracy







Results

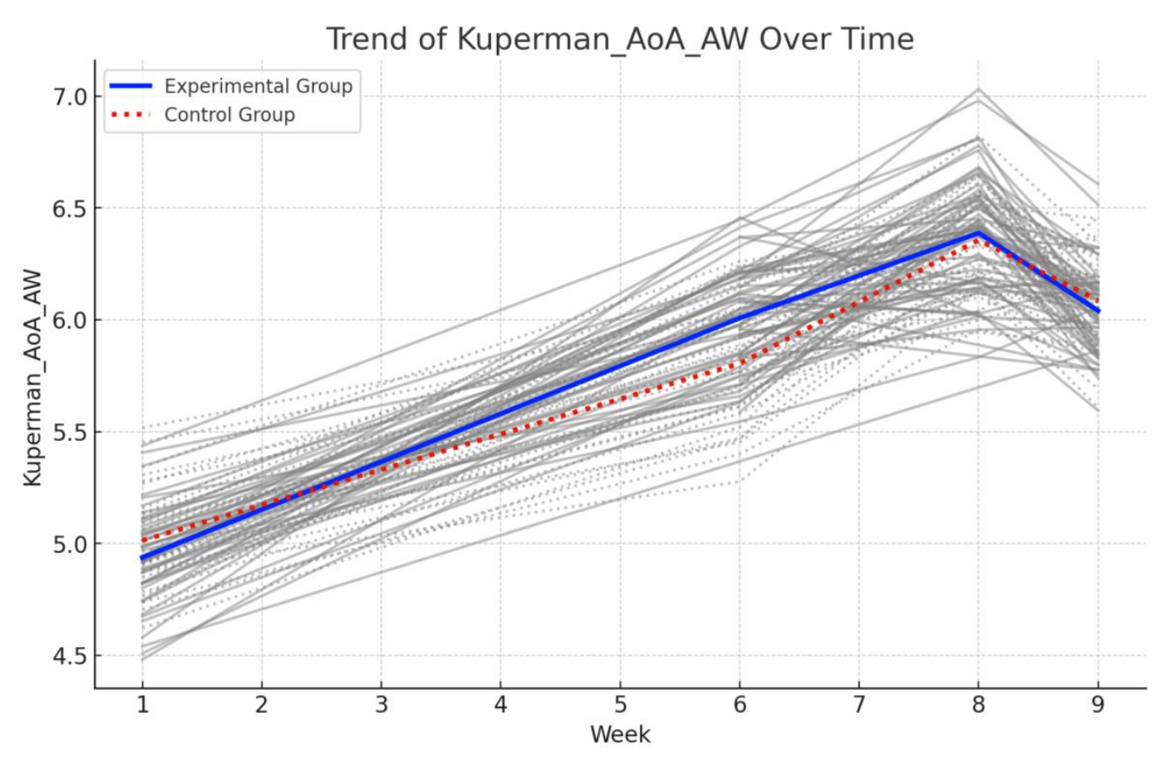
Syntactic complexity

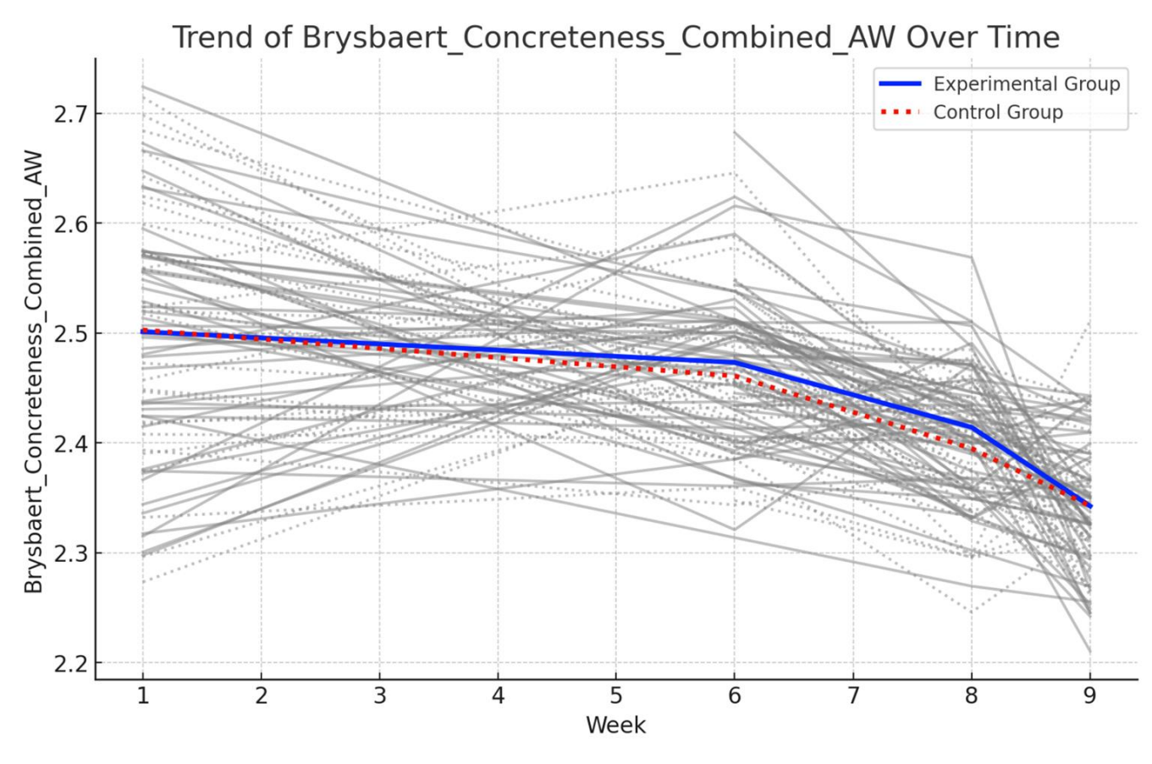
Effect of time (≈ instruction); some effect of experimental condition

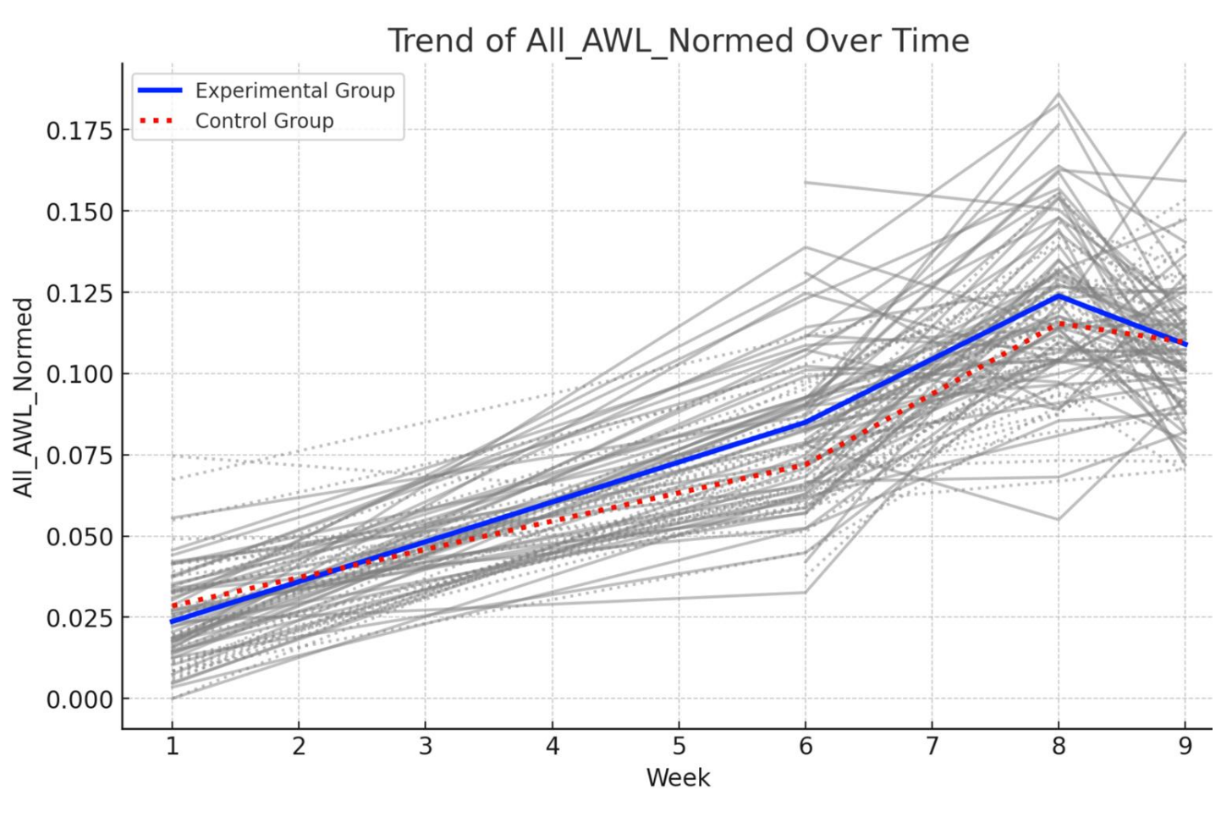
	Clauses / sent.		Complex nominals /Clause		Complex nominals / t-unit		Coordinated phrases / Clause		Coordinated phrases / t-unit		Dependent clauses / Clause		Dependent clauses / t-unit	
	Est.	*p*	Est.	*p*	Est.	*p*	Est.	*p*	Est.	*p*	Est.	*p*	Est.	*p*
(Intercept)	1.83	<.001	0.85	<.001	1.45	<.001	0.09	.006	0.15	0.001	0.54	<.001	0.92	<0.001
week	-0.08	**.001**	0.34	**<.001**	0.42	**<.001**	0.06	**<.001**	0.08	**<0.001**	-0.04	**<.001**	-0.09	**<0.001**
Condition	-0.03	.746	0.12	.264	0.01	.927	0.11	**.007**	0.14	**0.014**	-0.11	**.001**	-0.23	**0.010**
Week* condition	0.01	.669	0.00	.914	0.06	.215	-0.02	.203	-0.02	0.340	0.03	.005	0.07	0.01

Lexical complexity

Effect of time (≈ instruction); no differential impact of experimental condition







	Age of Aquisition						Concreteness						SUBTLEXus		Academic words	
	all		content		function		all		content		function		function, Log			
	Est.	p	Est.	p	Est.	p	Est.	p	Est.	p	Est.	p	Est.	p	Est.	p
Intercept	5.357	**<.001**	5.289	**<.001**	5.412	**<.001**	3.481	**<.001**	3.509	**<.001**	3.467	**<.001**	2.849	**<.001**	1.746	**<.001**
Week	-0.042	**.013**	-0.039	**.018**	-0.045	**.010**	0.019	**.040**	0.021	**.028**	0.018	**.045**	-0.027	**.023**	0.014	**.035**
Condition	0.028	.746	0.026	.264	0.032	.927	0.054	.927	0.049	.746	0.051	.746	-0.038	.927	0.017	.746
Week*Condition	0.012	.669	0.009	.914	0.014	.215	0.007	.215	0.008	.340	0.006	.669	-0.003	.203	0.006	.669

← download poster + references

GHENT
UNIVERSITY