Macrothink Institute™

International Journal of Linguistics
ISSN 1948-5425
2025, Vol. 17, No. 1

CALF Analysis of a Learner Corpus of English as a Second Language Students: Research Design Perspectives

Nickolas Komninos

Department of Languages and Literatures, Communication, Education and Society (DILL)

University of Udine, Italy

Palazzo Antonini, Via Petracco n. 8, Udine, Italy

Received: December 25, 2024 Accepted: January 20, 2025 Published: February 10, 2025

doi:10.5296/ijl.v17i1.22510 URL: https://doi.org/10.5296/ijl.v17i1.22510

Abstract

This study explores the application of CALF (Complexity, Accuracy, Lexical Complexity, and Fluency) measures in analysing the L2 writing. It replicates and expands on previous CALF research, addressing methodological and design considerations using the Swedish Learner English Corpus (SLEC). Using computational linguistics tools such as the L2 Syntactic Complexity Analyzer (L2SCA) and Coh-Metrix, the research evaluates texts from Swedish high school students across multiple dimensions. Through comparison with previous studies, it highlights genre-related differences and contextualises findings within pedagogical and linguistic theory frameworks. Despite limitations such as small sample size and motivational variances, the results contribute valuable insights into the interplay between CALF measures, L2 proficiency, and writing quality. Future research directions include refining learner corpus design and exploring the relationship between CALF measures and genre-specific writing quality.

Keywords: CALF measures, Learner corpus, Syntactic complexity, Lexical complexity, Writing quality, Second language acquisition, Computational linguistics, Genre analysis

1. Introduction

This paper uses a learner corpus and analyses it for syntactic complexity, accuracy, lexical complexity and fluency (CALF). This current study is placed within a corpus-based approach to writing research, specifically looking at pedagogic and research applications of CALF measures on monitoring L2 writing using the Swedish Learner English Corpus (SLEC) (Kaatari et al., 2024) to replicate previous CALF research. The use of learner corpora has also

found increasing traction within applied linguistic research (Granger, 2004; Granger et al, 2015; Meunier, 2020) and also specifically in writing research (Ädel, 2008). It is with this premise that the study uses learner corpora to monitor L2 writing with a view to improving research design.

This paper borrows from and extends previous syntactic complexity studies carried out by Lu (2011) and CALF studies carried out by Yoon and Polio (2017). Lu used a computational linguistics tool on a learner corpus to measure syntactic complexity. He also explored the use of this system to investigate other issues in L2 writing development and assessment linked with the impact of sampling conditions on the relationship between syntactic complexity and language development. Yoon and Polio replicate this on a learner corpus and extend the analysis from just syntactic complexity to include accuracy, lexical complexity and fluency (CALF). This study uses the methods of Yoon and Polio (2017) on a new learner corpus (SLEC) to investigate and extend research design of CALF analysis methods on L2 writing. Compared to Lu (2011), the use of CALF reduces the amount of data that can be analysed as the accuracy element of CALF is not automatically elaborated with computational linguistics tools, whereas syntactic and lexical complexity data are. The L2 Syntactic Complexity Analyzer (L2SCA) and Coh-Metrix computational linguistics tools are used to measure 15 complexity aspects in English as a Second Language (ESL) students' writing.

While the primary motivation behind this study was to address research design considerations, it also yields valuable implications for linguistic theory, particularly concerning the origins of genre-related differences, as well as pedagogical practices.

2. Methods

A random sample of L2 writing texts were taken from an existing corpus, the Swedish Learner English Corpus (SLEC) (Kaatari et al., 2024), which contains argumentative texts written by Swedish junior and senior high school students. SLEC provides rich metadata on the students' background, making it possible to empirically study relations between the linguistic properties of student texts and various extralinguistic and learner variables. A complete table of the metadata included in the corpus is listed in their 2024 publication. SLEC also includes a subset that has been assessed for second language (L2) proficiency. For proficiency, rather than adopting the 'proxies' approach of Yoon and Polio, a text-centred approach was used employing the Common European Framework of Reference for Languages (CEFR) scale, a common standard against which the assessment of language proficiency can be referenced.

2.1 Measures Used in CALF Analysis

- Syntactic complexity was assessed using the L2 Syntactic Complexity Analyzer (L2SCA) using 12 of the 14 measures offers by L2SCA (Lu, 2011).
- Lexical Complexity was measured with three indices from the Coh-Metrix computational tool (McNamara, Graesser, McCarthy, & Cai, 2014): average word length (WL), word frequency index (WF), and a lexical diversity measure (Vocd-D).

- Accuracy was assessed according to four error types (syntactic, morphological, preposition, and spelling errors per 100 words) were examined for accuracy. (Yoon & Polio, 2017)
- Fluency was measured by the number of words produced within a 30-minute time frame. (Yoon & Polio, 2017)

Table 1. Types of CALF Measures

Construct	Measures
Syntactic complexity	
Length of unit	Mean length of clause (MLC)
	Mean length of sentence (MLS)
	Mean length of T-unit (MLT)
Subordination	Clauses per T-unit (C/T)
	Dependent clauses per clause (DC/C)
	Dependent clauses per T-unit (DC/T)
Coordination	Coordinate phrases per clause (CP /C)
	Coordinate phrases per T-unit (CP /T)
	T-units per sentence (T/S)
Particular structures	Complex nominals per clause (CN/C)
	Complex nominals per T-unit (CN/T)
	Verb phrases per T-unit (VP /T)
Lexical complexity	Mean length of word (WL)
	Word frequency (WF)
	Vocd-D
Accuracy	Syntactic errors per 100 words
	Morphological errors per 100 words
	Preposition errors per 100 words
	Spelling errors per 100 words
Fluency	Total number of words in 30 minutes

10 randomly chosen files were taken from the SLEC corpus and analysed using the L2SCA and Coh-Metrix tools to create data for syntactic and lexical complexity measures, see Appendices A and B. The 10 samples were manually rated for accuracy according to the error-coding protocols described by Yoon and Polio (2017) in their Appendix B. An example of the analysis of the samples can be seen in Appendix C. Fluency was calculated according to word count as stated by L2SCA and times allocated for task included in the SLEC metadata. Table 2 illustrates the data.

Table 2. CALF Data

Sample no.		1	2	3	4	5	6	7	8	9	10	MEAN
Number of words of sample		880	572	399	315	224	501	371	213	754	423	
Syntactic Complexity												
Length of unit	(MLC)	8.89	9.08	7.53	9.84	6.22	7.37	7.42	6.87	7.32	8.13	**7.87**
	(MLS)	19.56	21.19	23.47	15.75	18.67	27.83	15.46	19.36	23.56	20.14	**20.50**
	(MLT)	16.92	18.45	16.63	15.00	18.67	25.05	13.25	16.38	18.39	16.92	**17.57**
Subordination	(C/T)	1.90	2.03	2.21	1.52	3.00	3.40	1.79	2.38	2.51	2.08	**2.28**
	(DC/C)	0.46	0.44	0.45	0.38	0.61	0.66	0.46	0.55	0.58	0.52	**0.51**
	(DC/T)	0.88	0.90	1.00	0.57	1.83	2.25	0.82	1.31	1.46	1.08	**1.21**
Coordination	(CP /C)	0.21	0.27	0.19	0.28	0.08	0.21	0.18	0.23	0.23	0.10	**0.20**
	(CP /T)	0.40	0.55	0.42	0.43	0.25	0.70	0.32	0.54	0.59	0.20	**0.44**
	(T/S)	1.16	1.15	1.41	1.05	1.00	1.11	1.17	1.18	1.28	1.19	**1.17**
Particular structures	(CN/C)	1.03	0.81	0.57	1.03	0.53	0.76	0.86	0.74	0.80	1.00	**0.81**
	(CN/T)	1.96	1.65	1.25	1.57	1.58	2.60	1.54	1.77	2.00	2.08	**1.80**
	(VP /T)	2.81	2.71	3.42	2.19	3.33	4.20	2.18	2.85	3.61	2.36	**2.97**
Lexical complexity	(WL)	4.31	3.92	4.09	4.04	3.87	4.16	3.93	3.85	3.95	3.90	**4.00**
	DESWLsy	1.40	1.31	1.32	1.29	1.29	1.35	1.25	1.28	1.27	1.25	**1.30**
	DESWLsyd	0.79	0.62	0.65	0.67	0.63	0.70	0.54	0.72	0.56	0.57	**0.65**
	DESWLlt	4.31	3.92	4.09	4.04	3.87	4.16	3.93	3.85	3.95	3.90	**4.00**
	DESWLltd	2.39	2.10	2.06	2.02	2.01	2.11	1.80	1.91	1.89	1.88	**2.02**
	(WF)	3.17	3.24	3.27	3.27	3.27	3.24	3.21	3.42	3.33	3.34	**3.27**
	Vocd-D	111.51	85.29	66.17	87.53	60.51	85.02	60.70	54.29	45.61	45.17	**70.18**
Accuracy	Syntactic	0.686	0.7	0.25	1.9	0	1.4	2.43	2.35	1.32	4.73	**1.58**
	Morphology	3.45	4.55	0.75	4.75	0.89	2.59	2.43	2.35	2.65	4.49	**2.89**
	Preposition	0	0.35	0	0.63	0	0.2	0	0	0.8	1.42	**0.34**
	Spelling	3.29	2.8	0.75	4.76	0	5.79	4.58	16.9	2.91	2.84	**4.46**
Fluency	Words/30n	440	286	199.5	157.5	112	250.5	160.5	106.5	377	211.5	**230.1**

As there were only 10 texts analysed, one off and multiple ANOVAs were not carried out.

3. Data Analysis and Results

Only argumentative texts were dealt with in the SLEC corpus, so no genre comparison was possible. There was also no native speaker group nor longitudinal elements, so learner development could not be monitored and native speaker comparison was impossible. It is interesting to take data from the Swedish corpus and compare it to Yoon and Polio's results.

Table 3 illustrates a data comparison of Yoon & Polio's three tests over time of the argumentative genre results, extracted from their appendix C together with the results from this replication:

Table 3. Comparison of replication data

Measure	Yoon and Polio Time 1 Mean (SD)	Yoon and Polio Time 2 Mean (SD)	Yoon and Polio Time 3 Mean (SD)	Replication Mean	Replication (SD)
MLC	8.42 (1.48)	8.23 (1.52)	8.49 (1.54)	7.87	1.11
MLS	17.41 (4.50)	17.11 (4.49)	17.58 (4.17)	20.50	3.74
MLT	15.34 (2.56)	14.85 (3.36)	15.06 (2.72)	17.57	3.11
C/T	1.85 (0.29)	1.83 (0.37)	1.80 (0.35)	2.28	0.57
DC/C	0.40 (0.08)	0.38 (0.11)	0.37 (0.10)	0.51	0.09
DC/T	0.75 (0.24)	0.72 (0.33)	0.70 (0.32)	1.21	0.51
CP/C	0.19 (0.09)	0.17 (0.10)	0.19 (0.12)	0.20	0.07
CP/T	0.34 (0.17)	0.29 (0.16)	0.35 (0.21)	0.44	0.16
T/S	1.13 (0.14)	1.15 (0.16)	1.17 (0.17)	1.17	0.11
CN/C	0.80 (0.23)	0.81 (0.25)	0.82 (0.24)	0.81	0.18
CN/T	1.45 (0.40)	1.47 (0.56)	1.45 (0.42)	1.80	0.38
VP/T	2.54 (0.43)	2.53 (0.71)	2.49 (0.48)	2.97	0.66
WL	4.44 (0.25)	4.50 (0.29)	4.46 (0.27)	4.00	0.15
WF	3.09 (0.07)	3.09 (0.10)	3.09 (0.11)	3.27	0.07
Vocd-D	68.94 (15.02)	66.81 (12.66)	69.32 (18.25)	70.18	21.44
SyntE	2.30 (1.20)	2.39 (1.59)	1.96 (1.24)	1.58	1.39
MorphE	7.40 (2.91)	7.23 (2.98)	7.29 (2.92)	2.89	1.43
PrepE	1.41 (0.93)	1.41 (1.01)	1.85 (1.14)	0.34	0.48
SpellE	2.93 (2.85)	2.39 (2.23)	1.68 (1.56)	4.46	4.71
Total W	294.16 (62.54)	307.70 (67.38)	321.16 (73.73)	230.1	110.45

As can be seen from the table, the replicated results reflect many of the original results, notwithstanding the difference in age, schooling, L1, prompt differences, institutional conditions and sample size. In one way this is a good thing as the researcher carrying out the study has been reassured that the calculations are comparable to the original ones. On the other hand, there are enough differences to have clear markers of separation between the two studies. For example, the accuracy results differ markedly, especially morphological error. This confirms Yoon & Polio's suggestion that accuracy and fluency results need to considered with caution and that more research in this area is needed. It would be worth examining how the SLEC metadata available for the randomised Swedish samples corresponded in terms of CALF results, for example for proficiency according to the CEFR, age, L1 etc., however this metadata is not available from the 2017 study.

It should also be noted that samples 3 and 5 of the randomised replication seem to be outliers with exceptionally few errors in accuracy. In larger cohorts, outliers like these have little

effect, but when there are only 10 samples, the effect is great. Good practice with smaller corpora would be to omit such outlying samples or apply a formula to account for the variation in results due to this. Any future learner corpus design should also consider how to deal with outlier results, also from a research ethics perspective.

The motivation that the participants perceive to write well has a significant effect on the quality of their writing, especially spelling. It would seem that some participants had little motivation to write well and to construct and represent an argument (see sample 7 and 8) and some were more motivated (sample 1, 2, 3, 5 and 9). The motivation factor is important to quantify and will have a huge impact on results. This type of task contextualisation is needed to make results more comparable.

Maybe the most important outcome from this replication actually came about during the accuracy analysis. It was noted that even when participants achieved very good values in the analysis, for example sample 3 and 5, their writing quality was not necessarily particularly good. Yoon and Polio also state as much. This highlights the focus of this type of research on the cohesive elements in text construction. Other aspects also contribute to good quality writing, not least coherence to a generic pattern, extent of use within the generic structure potential of a specific genre and also motivation to write well. For example, some of the texts analysed were fine cohesively, scored well in accuracy and the syntactical and lexical complexity aspects reflected those of other ESL and also native speakers but did not satisfy genre needs and were relatively incoherent to the functional communicative aim of the communication act. With this in mind, it was interesting to compare Yoon & Polio's native-speaker and non-native speaker results with non-native speaker replication results, see Table 4.

Table 4. Comparison of with native speaker results

	Time 1 Arg. Mean (SD)	Time 2 Arg. Mean (SD)	Time 3 Arg. Mean (SD)	Native speaker Argumentative Mean (SD)	Replication Argumentative Mean
Length of unit					
MLC	8.42 (1.48)	8.23 (1.52)	8.49 (1.54)	9.25 (1.44)	7.87
MLS	17.41 (4.50)	17.11 (4.49)	17.58 (4.17)	19.93 (3.74)	20.50
MLT	15.34 (2.56)	14.85 (3.36)	15.06 (2. 72)	17.55 (3.33)	17.57
Subordination					
C/T	1.85 (0.29)	1.83 (0.37)	1.80 (0.35)	1.92 (0.38)	2.28
DC/C	0.40 (0.08)	0.38 (0.11)	0.37 (0.10)	0.39 (0.11)	0.51
DC/T	0.75 (0.24)	0.72 (0.33)	0.70 (0.32)	0.78 (0.34)	1.21
Coordination					
CP/C	0.19 (0.09)	0.17 (0.10)	0.19 (0.12)	0.24 (0.11)	0.20
CP/T	0.34 (0.17)	0.29 (0.16)	0.35 (0.21)	0.45 (0.20)	0.44
T/S	1.13 (0.14)	1.15 (0.16)	1.17 (0.17)	1.14 (0.11)	1.17
Particular structure					

CN/C	0.80 (0.23)	0.81 (0.25)	0.82 (0.24)	0.90 (0.24)	0.81
CN/T	1.45 (0.40)	1.47 (0.56)	1.45 (0.42)	1.69 (0.43)	1.80
VP/T	2.54 (0.43)	2.53 (0.71)	2.49 (0.48)	2.65 (0.53)	2.97
Lexical complexity					
WL	4.44 (0.25)	4.50 (0.29)	4.46 (0.27)	4.74 (0.22)	4.00
WF	3.09 (0.07)	3.09 (0.10)	3.09 (0.11)	3.05 (0.08)	3.27
Vocd-D	68.94 (15.02)	66.81 (12.66)	69.32 (18.25)	84.78 (15.52)	70.18
Accuracy					
SyntE	2.30 (1.20)	2.39 (1.59)	1.96 (1.24)		1.58
MorphE	7.40 (2.91)	7.23 (2.98)	7.29 (2.92)		2.89
PrepE	1.41 (0.93)	1.41 (1.01)	1.85 (1.14)		0.34
SpellE	2.93 (2.85)	2.39 (2.23)	1.68 (1.56)		4.46
Fluency					
Total W	294.16 (62.54)	307.70 (67.38)	321.16 (73.73)	333.75 (98.33)	230.1

Some observations can be made on the basis of these results. The SLEC replication sample had MLS and MLT closer to the native-speaker cohort, but MLC lower than both the cohorts in Yoon & Polio's results. Lu 2011 states that length-of-unit measures discriminated between his levels one and two, and complex nominals between his one and two, and two and three. CN/T results for the SLEC cohort is also high, higher than the native-speaker cohort, but CN/C is equivalent to Yoon & Polio's NNS cohort. The SLEC cohort showed more instances of subordination than the other cohorts, and more instances of coordination than the ESL group, equivalent to the NS group. Bulté and Housen (2014) suggest that mean length of sentence MLS and MLT corresponded to both development and quality, whereas measures related to the use of dependent clauses were related to quality. The SLEC cohort also showed shorter word lengths than the other cohorts, but more variety of vocabulary than the ESL cohort, and less than the NS cohort.

Interestingly, but to be considered with caution, are the accuracy and fluency results, which show a significantly lower morphological error rate for the SLEC group and a far higher spelling error rate and far fewer words produced. It has been hypothesised that motivational factors could account for the high spelling error rate and less text production in the SLEC cohort, although other contextualised factors could also account for this, like L2 proficiency or age. It would also be interesting to compare these CALF result to assessments of quality of the specific texts, which is beyond the aims of these studies, but could be an avenue of future research.

4. Discussion

This replication of Yoon & Polio's study has demonstrated that there is a lot of information about genre and L2 development that can potentially be extracted and monitored through learner corpora and CALF measures. It suggests that functional differences in communication needs between genres better explains variations in language complexity. The native speakers

of the 2017 study showed the same pattern of genre effects as the L2 writers with regard to increased length-of-unit complexity in argumentative writing and the lack of genre differences in the clause-level measures. The lexical results were also the same, both groups used longer and less frequent words in the argumentative genre and a greater diversity of words in the narratives. They would have expected to see accuracy differences if the cognition hypothesis (Robinson; 2001; 2005) or limited attentional capacity model (Skehan, 1998) applied to genre differences, but they did not. The study examines lexical measures and attributes the increased length of words in argumentative essays to the greater use of nominalisation, (Byrnes et al., 2010) leading to longer average word length. It also notes lower lexical variety in argumentative essays, potentially due to formulaic phrases commonly used in this genre.

One notable finding is the lack of genre effects on the clause-level measures, as opposed to the phrasal measures. This may be because these texts include some distinctive features of academic texts, this may corroborate the argument that academic writing is characterised by the increased use of complex phrases, but not by clause-level sophistication (Biber, Gray, & Poonpon, 2011; Byrnes, Maxim, & Norris, 2010). These results emphasise the importance of controlling for genre when conducting research on L2 writing development and intervention effects. It suggests that students at all proficiency levels should be exposed to a variety of genres in their writing assignments to encourage the use of more complex language. The study reinforces that while language complexity may vary across genres, accuracy may not show significant changes, emphasising that errors are a natural part of the language learning process. It raises questions about the potential benefits of having students write in genres that elicit more complex language over time and calls for further research in this area.

Learner corpora design and the metadata that contextualises the communication event being recorded is key to having comparable data, both between cohorts, within one cohort over time and even between different participants of a specific corpus. To adequately inform genre studies, more research is still required on the relationship between CALF measures and quality of writing. Defining what is good writing in one genre as opposed to another, with CALF measure but also with other measures, could be a good start. With this in mind, research could focus on investigating which aspects contribute to good quality writing in specific genres, be it cohesive elements, coherence, the use of rhetorical devices, organisation of information, motivation to use a large percentage of generic structure potential of a genre and so on. CALF measures can certainly contribute to that knowledge.

When carrying out a longitudinal learner corpus study, there are many pressures that influence choices made in that research activity. The attrition between the quantity of result and quality of results is apparent in almost all research of this type. The points below are simply intended to direct future research to improve the research design and methods, and are not intended to detract from any of the important findings that have been achieved in previous studies that have dealt with this area of research. Firstly, proficiency levels need to be well defined and replicable. Following a text-centred levels system (Gilquin, 2015), like the Common European Framework of Reference for languages. This was not the case before

SLEC, which deals with this using the CEFR in its metadata and this can be seen as an advancement in learner corpora metadata collection.

Cohorts are small due to the need to hand-rate the accuracy aspect of CALF. This stops research results from having the same large-scale data input that informed Lu's 2011 study. As accuracy and fluency are also to be viewed with caution due to a lack of research in the area and non-robust nature of its statistics, maybe accuracy and fluency should be dealt with separately in future studies, thus separating syntactic and lexical complexity from accuracy and fluency. A lack of differences using inferential statistics cannot be considered a robust finding because such statistics are not designed to detect a lack of difference (e.g., Godfroid & Spino, 2015). Also, linguistic development in L2 writing is complex and influenced by factors like proficiency levels and task control.

Comparing non-native speaker with native speaker is a good research method but there is a weakness in the design that does not allow the full potential of the research and data to emerge. All ESL non-native speakers are often grouped together. As was done by SLEC, it would be more informative and a better research design to create a distinction and data-traceability also according to L1, language spoken at home and so on. In this way, observations could be made on how different L1 influences L2 writing. This would be interesting not only from a research point of view, but could also inform pedagogy as also remedial strategies could be identified based on the results of this research.

Longitudinal studies often only last 1 or 2 semesters, this is not enough time to meaningfully investigate development in writing. Intervals over a longer period of time would produce more relevant results, but of course there are research pressures that do not allow for this kind of longitudinal study. Native speaker cohorts often only look at some measures, omitting data on accuracy. This is a gap in the results that is probably due to the time it takes to manually rate the texts. For fuller research products this gap also needs to be addressed.

The impact of sample size needs to be highlighted, which was mentioned in Lu's research (2011) to explain discrepancies with previous research on syntactic complexity. Yoon and Polio discuss the use Lu's use of multiple ANOVAs in relation to this aspect but anyway feel that their results confirm Lu's findings. The disproportionate influence of outliers on small corpus studies must be dealt with in a way that makes separate studies comparable.

Description of the motivational factors for participants to write well must be incorporated and standardised. This could have an impact on the care and attention the participants have in carrying out the task. These last two points are part of a bigger problem about what metadata to include in learner corpora. Previously to SLEC not much metadata nor contextualisation was given in reference to the learner corpus. SLEC has approached this, especially in the area of extramural English activities, however it would be good best practice to include more metadata on motivational factors and contextualisation in future learner corpora, thus opening those corpora to wider avenues of research.

BAAL describe some ethical issues in the building corpora dealing with anonymity and informed consent. According to Lancaster University, ethical considerations are rarely found

in corpus linguistics but do detail issues relating to respondents, distributors and users of corpora data. The SLEC and data used in this current study follow all BAAL and Lancaster University guidelines by maintaining anonymity. All texts included in SLEC were pseudonymised to make sure that there is no personal information included which could be used to identify participants. All students were informed that participation in SLEC is voluntary and that their texts would be pseudonymised. Parental consent had been collected for students under the age of 15. Any future learner corpora construction that collects detailed metadata to aid research should also follow these ethical guidelines.

This study reaffirms the value of genre-specific research in understanding L2 writing development and highlights the limitations of existing learner corpora in capturing the full complexity of this process. Addressing issues such as small sample sizes, insufficient longitudinal scope, and inadequate metadata will be crucial for advancing the field. With improved corpus design and a focus on CALF measures alongside broader evaluative criteria, future studies can provide deeper insights into what constitutes quality writing across genres and how best to support L2 writers in achieving it.

Acknowledgments

I would like to express my sincere gratitude to Henrik Kaatari, Tove Larsson, and Ying Wang for their generous permission to use the Swedish Learner English Corpus (SLEC). Access to such learner corpora is invaluable, and this research would not have been possible without their support. I am also deeply grateful to Joseph Lee for his valuable advice and guidance.

References

Ädel, A. (2008). Involvement features in writing: Do time and interaction trump register awareness?. In G. Gilquin, S. Papp, & M. B. D ęz-Bedmar (Eds.), *Linking up contrastive and learner corpus research* (pp. 35-53). Amsterdam: Rodopi. https://doi.org/10.1163/9789401206204_003

Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development?. *Tesol Quarterly*, *45*(1), 5-35. https://doi.org/10.5054/tq.2011.244483

Bulté, B., & Housen, A. (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing*, *26*, 42-65. https://doi.org/10.1016/j.jslw.2014.09.005

Byrnes, H., Maxim, H. H., & Norris, J. M. (2010). Realizing advanced foreign language writing development in collegiate education: Curricular design, pedagogy, assessment. *The Modern Language Journal*, i-235. Retrieved from https://doi-org.www.bibproxy.du.se/10.1111/j.1540-4781.2010.01136.x

Godfroid, A., & Spino, L. A. (2015). Reconceptualizing reactivity of think‐alouds and eye tracking: Absence of evidence is not evidence of absence. *Language Learning*, *65*(4), 896-928. https://doi.org/10.1111/lang.12136

Granger, S. (2004). Computer learner corpus research: Current status and future prospects. *Applied Corpus Linguistics*, 123-145. https://doi.org/10.1163/9789004333772_008

Granger, S., Gilquin, G., & Meunier, F. (Eds.) (2015). *The Cambridge handbook of learner corpus research.* Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139649414

Kaatari, H., Wang, Y., & Larsson, T. (2024). Introducing the Swedish Learner English Corpus: A corpus that enables investigations of the impact of extramural activities on L2 writing. *Corpora*, *19*(1). https://doi.org/10.3366/cor.2024.0296

Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly*, *45*(1), 36-62. Retrieved from https://doi-org.www.bibproxy.du.se/10.5054/tq.2011.240859

McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press. https://doi.org/10.1017/CBO9780511894664

Meunier, F. (2020). Introduction to learner Corpus research. In *The Routledge handbook of second language acquisition and corpora* (pp. 23-36). Routledge. https://doi.org/10.4324/9781351137904-4

Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics*, *22*(1), 27-57. https://doi.org/10.1093/applin/22.1.27

Robinson, P. (2005). Cognitive complexity and task sequencing: Studies in a componential framework for second language task design. https://doi.org/10.1515/iral.2005.43.1.1

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press. https://doi.org/10.1177/003368829802900209

Yoon, H. J., & Polio, C. (2017). The linguistic development of students of English as a second language in two written genres. *Tesol Quarterly*, *51*(2), 275-301. https://doi.org/10.1002/tesq.296

Appendix A. L2SCA data of 10 samples from SLEC

	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6	Sample 7	Sample 8	Sample 9	Sample 10
Measure										
nwords	880.00	572.00	399.00	315.00	224.00	501.00	371.00	213.00	754.00	423.00
MLS	19.56	21.19	23.47	15.75	18.67	27.83	15.46	19.36	23.56	20.14
MLT	16.92	18.45	16.63	15.00	18.67	25.05	13.25	16.38	18.39	16.92
MLC	8.89	9.08	7.53	9.84	6.22	7.37	7.42	6.87	7.32	8.13
C_S	2.20	2.33	3.12	1.60	3.00	3.78	2.08	2.82	3.22	2.48
VP_T	2.81	2.71	3.42	2.19	3.33	4.20	2.18	2.85	3.61	2.36
C_T	1.90	2.03	2.21	1.52	3.00	3.40	1.79	2.38	2.51	2.08
DC_C	0.46	0.44	0.45	0.38	0.61	0.66	0.46	0.55	0.58	0.52
DC_T	0.88	0.90	1.00	0.57	1.83	2.25	0.82	1.31	1.46	1.08
T_S	1.16	1.15	1.41	1.05	1.00	1.11	1.17	1.18	1.28	1.19
CT_T	0.50	0.52	0.67	0.38	0.75	0.80	0.64	0.69	0.66	0.60
CP_T	0.40	0.55	0.42	0.43	0.25	0.70	0.32	0.54	0.59	0.20
CP_C	0.21	0.27	0.19	0.28	0.08	0.21	0.18	0.23	0.23	0.10
CN_T	1.96	1.65	1.25	1.57	1.58	2.60	1.54	1.77	2.00	2.08
CN_C	1.03	0.81	0.57	1.03	0.53	0.76	0.86	0.74	0.80	1.00

Appendix B. Coh-Metrix Data of 10 Samples from SLEC

Sample	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
DESPC	6	7	7	6	6	6	8	6	6	7
DESSC	46	26	18	21	13	19	26	11	32	23
DESWC	869	572	399	311	224	491	373	211	754	423
DESPL	7.667	3.714	2.571	3.5	2.167	3.167	3.25	1.833	5.333	3.286
DESPLd	5.502	1.799	1.272	2.345	0.753	1.169	1.488	1.602	2.944	2.563
DESSL	19.043	22.115	22.278	14.81	17.692	25.842	14.346	19.182	24	18.783
DESSLd	9.29	12.545	7.307	8.322	8.92	14.523	6.449	7.744	12.384	7.354
DESWLsy	1.405	1.309	1.318	1.289	1.286	1.35	1.249	1.28	1.269	1.251
DESWLsyd	0.793	0.621	0.655	0.672	0.634	0.699	0.543	0.719	0.563	0.567
DESWLlt	4.306	3.916	4.09	4.042	3.866	4.165	3.933	3.848	3.952	3.903
DESWLltd	2.394	2.097	2.055	2.018	2.009	2.108	1.797	1.909	1.891	1.881
PCNARz	1.047	1.307	2.129	0.845	2.217	1.538	1.395	1.473	1.684	1.431
PCNARp	85.08	90.32	98.3	79.95	98.64	93.7	91.77	92.92	95.35	92.36
PCSYNz	-0.495	-0.789	-1.159	0.187	-0.517	-0.764	-0.146	-0.967	-0.829	-0.802
PCSYNp	31.21	21.48	12.51	57.14	30.5	22.36	44.43	16.85	20.33	21.19
PCCNCz	-0.865	0.821	-0.306	-0.462	-0.639	-0.737	0.244	-1.599	-0.438	-1.666
PCCNCp	19.49	79.39	38.21	32.28	26.11	23.27	59.48	5.59	33.36	4.85

PCREFz	-0.778	0.712	1.797	-1.104	1.67	0.94	1.534	1.771	2.422	2.131
PCREFp	22.06	76.11	96.33	13.57	95.15	82.64	93.7	96.16	99.22	98.34
PCDCz	1.206	2.553	1.828	0.571	3.309	0.497	1.339	0.113	1.074	4.149
PCDCp	88.49	99.46	96.56	71.57	99.95	68.79	90.82	54.38	85.77	100
PCVERBz	0.72	1.394	1.162	1.201	1.232	0.467	1.307	1.989	2.165	2.526
PCVERBp	76.42	91.77	87.7	88.49	89.07	67.72	90.32	97.61	98.46	99.41
PCCONNz	-2.648	-5.417	-2.955	-3.671	-3.42	-3.475	-3.178	-4.049	-3.573	-2.922
PCCONNp	0.41	0	0.16	0	0.03	0.03	0.08	0	0	0.18
PCTEMPz	0.159	-0.128	-0.643	-0.015	0.959	0.315	0.048	0.542	0.095	1.267
PCTEMPp	55.96	45.22	26.11	49.6	82.89	62.17	51.6	70.54	53.59	89.62
CRFNO1	0.244	0.44	0.471	0.15	0.5	0.5	0.48	0.6	0.419	0.591
CRFAO1	0.489	0.72	0.941	0.35	0.75	0.722	0.8	0.8	0.806	0.727
CRFSO1	0.244	0.48	0.529	0.2	0.583	0.778	0.52	0.7	0.484	0.591
CRFNOa	0.158	0.234	0.248	0.11	0.24	0.378	0.39	0.6	0.332	0.451
CRFAOa	0.417	0.541	0.904	0.239	0.587	0.563	0.61	0.855	0.777	0.703
CRFSOa	0.21	0.263	0.352	0.129	0.28	0.511	0.405	0.636	0.358	0.486
CRFCWO1	0.099	0.173	0.222	0.073	0.264	0.164	0.218	0.187	0.263	0.263
CRFCWO1d	0.115	0.187	0.082	0.113	0.172	0.144	0.155	0.094	0.166	0.119
CRFCWOa	0.065	0.102	0.196	0.06	0.163	0.115	0.156	0.21	0.222	0.226
CRFCWOad	0.081	0.115	0.107	0.097	0.155	0.114	0.157	0.121	0.162	0.155
CRFANP1	0.489	0.64	1	0.3	0.417	0.5	0.64	0.5	0.774	0.409
CRFANPa	0.17	0.229	0.496	0.077	0.36	0.252	0.141	0.564	0.543	0.423
LSASS1	0.095	0.241	0.164	0.13	0.35	0.309	0.235	0.258	0.273	0.254
LSASS1d	0.086	0.196	0.089	0.087	0.209	0.147	0.187	0.143	0.209	0.149
LSASSp	0.081	0.196	0.185	0.095	0.319	0.257	0.267	0.301	0.225	0.254
LSASSpd	0.087	0.143	0.086	0.104	0.197	0.161	0.195	0.127	0.179	0.145
LSAPP1	0.345	0.352	0.293	0.215	0.396	0.424	0.277	0.152	0.488	0.38
LSAPP1d	0.146	0.151	0.108	0.122	0.146	0.094	0.201	0.061	0.139	0.054
LSAGN	0.264	0.34	0.308	0.24	0.382	0.361	0.357	0.329	0.419	0.369
LSAGNd	0.072	0.108	0.114	0.121	0.147	0.112	0.132	0.136	0.12	0.117
LDTTRc	0.71	0.591	0.622	0.693	0.586	0.569	0.545	0.618	0.342	0.428
LDTTRa	0.432	0.39	0.406	0.492	0.43	0.377	0.381	0.479	0.212	0.299
LDMTLD	97.407	71.003	62.635	83.497	48.402	69.297	44.444	54.072	45.333	40.004
LDVOCD	111.513	85.289	66.172	87.534	60.511	85.017	60.7	54.289	45.612	45.173
CNCAll	105.869	125.874	115.288	102.894	133.929	109.98	109.92	94.787	116.711	115.839
CNCCaus	26.467	33.217	37.594	19.293	58.036	16.293	42.895	18.957	22.546	54.374
CNCLogic	43.728	82.168	50.125	51.447	66.964	46.843	42.895	47.393	71.618	89.835
CNCADC	16.11	41.958	12.531	16.077	26.786	18.33	16.086	23.697	21.22	18.913
CNCTemp	26.467	12.238	22.556	25.723	13.393	16.293	10.724	4.739	14.589	9.456
CNCTempx	14.96	10.49	12.531	12.862	4.464	6.11	13.405	14.218	9.284	16.548
CNCAdd	55.236	76.923	62.657	64.309	53.571	63.136	61.662	66.351	63.66	54.374
CNCPos	97.814	96.154	105.263	96.463	116.071	91.65	107.239	71.09	96.817	111.111

CNCNeg	10.357	33.217	12.531	12.862	22.321	16.293	8.043	23.697	19.894	14.184
SMCAUSv	26.467	17.483	20.05	38.585	26.786	24.44	40.214	23.697	19.894	18.913
SMCAUSvp	36.824	41.958	47.619	45.016	71.429	30.55	67.024	23.697	34.483	54.374
SMINTEp	6.904	5.245	15.038	16.077	8.929	4.073	10.724	0	13.263	2.364
SMCAUSr	0.375	1.273	1.222	0.154	1.429	0.231	0.625	0	0.688	1.667
SMINTEr	2.714	3	1.714	1	3.667	2.667	2.4	4	0.727	7.5
SMCAUSlsa	0.048	0.082	0.088	0.07	0.094	0.074	0.073	0.14	0.125	0.177
SMCAUSwn	0.459	0.508	0.442	0.429	0.447	0.367	0.529	0.483	0.661	0.4
SMTEMP	0.867	0.82	0.794	0.85	0.917	0.889	0.86	0.9	0.855	0.909
SYNLE	3.957	4.615	3.611	2.429	5.615	2.316	3.308	2.455	3.938	3.043
SYNNP	0.65	0.647	0.409	0.727	0.433	0.579	0.721	0.544	0.546	0.802
SYNMEDpos	0.688	0.646	0.654	0.715	0.674	0.662	0.639	0.665	0.628	0.62
SYNMEDwrd	0.908	0.872	0.848	0.941	0.894	0.882	0.862	0.867	0.815	0.818
SYNMEDlem	0.895	0.858	0.822	0.937	0.857	0.833	0.84	0.804	0.785	0.803
SYNSTRUTa	0.071	0.082	0.046	0.065	0.063	0.056	0.088	0.064	0.065	0.068
SYNSTRUTt	0.065	0.055	0.057	0.068	0.057	0.047	0.099	0.084	0.064	0.06
DRNP	337.169	333.916	333.333	347.267	308.036	362.525	359.249	312.796	327.586	321.513
DRVP	253.165	248.252	350.877	250.804	281.25	234.216	227.882	270.142	319.629	191.489
DRAP	32.221	43.706	42.607	38.585	40.179	54.99	34.853	18.957	33.156	42.553
DRPP	95.512	90.909	65.163	90.032	53.571	75.356	67.024	71.09	62.334	73.286
DRPVAL	0	0	5.013	6.431	4.464	4.073	2.681	9.479	0	0
DRNEG	19.563	8.741	15.038	25.723	31.25	16.293	5.362	37.915	21.22	28.369
DRGERUND	33.372	13.986	32.581	12.862	8.929	0	5.362	9.479	30.504	0
DRINF	31.07	31.469	42.607	32.154	17.857	18.33	10.724	23.697	43.767	4.728
WRDNOUN	189.874	206.294	145.363	202.572	138.393	207.739	198.392	194.313	151.195	182.033
WRDVERB	147.296	111.888	142.857	102.894	111.607	101.832	99.195	99.527	95.491	82.742
WRDADJ	79.401	66.435	65.163	90.032	75.893	61.1	96.515	123.222	86.208	113.475
WRDADV	64.441	78.671	75.188	90.032	98.214	89.613	64.343	66.35	64.986	104.019
WRDPRO	105.869	106.643	147.87	106.109	147.321	136.456	166.22	104.265	140.584	113.475
WRDPRP1s	34.522	52.448	15.038	12.862	62.5	20.367	75.067	37.915	19.894	18.913
WRDPRP1p	1.151	1.748	0	0	0	26.477	10.724	0	1.326	0
WRDPRP2	33.372	17.483	92.732	54.662	75.893	61.1	48.257	37.915	102.122	87.47
WRDPRP3s	0	12.238	0	0	0	0	10.724	0	0	0
WRDPRP3p	14.96	5.245	7.519	25.723	0	22.403	5.362	18.957	1.326	2.364
WRDFRQc	2.575	2.678	2.741	2.671	2.844	2.741	2.75	3.017	2.822	2.872
WRDFRQa	3.167	3.243	3.265	3.269	3.27	3.237	3.208	3.417	3.326	3.337
WRDFRQmc	1.511	1.651	1.496	1.529	2.202	1.96	1.69	1.377	1.316	1.602
WRDAOAc	348.618	296.911	325.5	332.476	308.478	331.545	325.091	354.667	339.171	331.074
WRDFAMc	584.032	591.928	591.099	586.258	598.194	590.273	590.115	587.293	591.926	596.505
WRDCNCc	335.121	356.883	329.076	356.237	335.052	327.068	351.692	310.531	334.798	317.045
WRDIMGc	372.896	408.738	371.583	393.694	399.265	384.465	405.402	362.537	381.047	370.342
WRDMEAc	418.781	457.155	433.221	433.667	448.034	437.468	454.237	426.257	444.885	430.166

WRDPOLc	4.587	5.047	5.603	4.417	5.63	4.206	5.409	4.996	5.197	5.108
WRDHYPn	6.051	5.884	5.579	5.415	6.367	4.743	6.21	4.592	5.857	6.197
WRDHYPv	1.449	1.349	1.375	1.368	1.46	1.393	1.509	1.26	1.205	1.363
WRDHYPnv	1.357	1.39	1.128	1.25	1.182	1.227	1.411	1.092	1.12	1.266
RDFRE	68.798	73.764	72.833	82.753	80.55	66.395	86.608	79.077	75.561	82.334
RDFKGL	8.356	8.436	8.608	5.396	6.305	10.418	4.743	6.995	8.574	6.344
RDL2	21.673	28.489	30.253	22.072	35.778	27.836	32.816	35.653	35.351	36.64

Appendix C. Example sample from SLEC analysed for accuracy

1

The fundamentals for a good life

Living a good life might be one of the hardest things to achieve. There are so many factors, internal and external that can push your life in an unlimited amount of directions. This leads to life being quite unpredictable which stresses a lot of people out. From people my age I hear a lot about wanting to improve as a person, study, eat healthy and becoming what the internet has named ""that girl"". But at the same time it feels like a lot of people base thier personality in lazyness and despair. Turning mental health problems, small and large, into an aesthetic. I believe that there is no way to live a perfect life with all of its ups and downs. I think that finding the balance between self care and personal growth while also leaving room for downfalls and procrastination are the most important aspects for leading a good life.

We've all heard the classic ""eat, sleep and workout"" advice countless times. Guidence counselors, coaches, teachers, parents, they all say the same thing. But when put into reality, following this mold can be very difficult. Sure, they work as general guidelines for eventual better habits, but they don't win any efficency awards. How am I supposed to stay motivated an on my game 24/7 when I have a million other things going on in my life at once? In the end it all comes down to prioritizing, and a not lot of people are willing to change their lifestyle for vague and slowly growing results. Letting go of your responsibilities and just letting yourself breath for a second is so important because it allows you to reflect on your natural habits. Things such as what you do when you feel overwhelmed or how you react to stress are crucial to take note of in order to even try to make a change. When things get too much it's okay to take a break, it's okay to push tasks to the last second or not even doing them at all, if it meant that you got some time for yourself. What you choose to do with that time is not something that I should interfere with, in fact no one has the right to say what you should do to let your brain rest. In situations like this it is of outmost ungency to put your mental health first.

SyntE 6
Morph 19
Prep 0
Spell 29

Words 880

The thing i've noticed with all of these lifestyle gurus and influencers is that they seem to forget that not everyone are as motivated and diciplined as they are. My guess is that most of them don't even follow all the advice that they keep spitting out onto their followers. Because after all, they are getting payed for it, and no one knows what goes on behind the camera. But what they do good, is that they provide inspiration. I would have never started planning and writing things in my bullet journal it wasn't for the internet providing me with information on how to compleatly change my life with an overpriced notebook. Did it change my life? No. Is it making me more organized and helps me plan out my time? Yes. This is the balance that I mentioned earlier. This perfct lifestyle that you encounter on pinterest boards and instagram posts is not something that anyone is expecting you to acheive. No one actually lives like that. But using those stupidly saturated photos as inspiration and motivation to improve an aspect of your life is something I highly recomend. Learn, watch, write and take in as much information as your heart desires, but make sure to remember to keep your expectations resonable. Only put into action what you truly believe will work for you.

One might read this and think that I'm some sort of pessemistic freak who encourages people to see everything with a doubtfull eye.But I belive that I'm just being realistic. Don't come at me with your toxic positivity where everything is sunshine and rainbows and tell me that that is a realistic way to percieve life. Life isn't fair, it's not always fun and it somehow never goes as you plan. I like to compare it to a rollercoaster, it goes up and then you fall. It takes a sharp turn only to stop abruptly and all of a sudden your upside down. Imagine going trough all stages of life on the same terms, imagine a rollercoaster that jsut goes on and on on a flat rail. That doesn't sound very exciting does it? Because whats life without a little challenge? The feeling of recalling the past and thinking ""wow, I got trough that"" is, in my opinion, hard to match. Looking back and seeing where and who you are now as opposed to then is lifes own reciept that things will get better, and that life will go on.

With that, I belive that trying to do your best while also leaving room for your worst and keeping a realistic view on your possibillities are the most important aspects of leading a good life.