# Linguistic Complexity as an Indicator of Writing Quality

School of Computer Science and Statistics, Trinity College Dublin
Oisín Nolan • oinolan@tcd.ie

## 1 Introduction

The development of writing skills is essential to success in many areas, both in education and the workplace (McNamara et al., 2010). Despite its importance, these skills are slow to develop, and generally of a low standard (Ferretti and Graham, 2019). This work explores the relationship between linguistic complexity and writing quality in a corpus of 1782 persuasive essays written by eighth-grade students. More specifically, the study aims to identify correlates of awarded essay grade among various measures of syntactic and lexical complexity. Understanding the linguistic features that correlate with writing quality could find application in areas such as personalised pedagogical feedback and the development of automated essay scoring systems (McNamara et al., 2014; Kumar and Boulanger, 2020). With these applications in mind, the set of essays was partitioned into 3 groups: *high*, *medium*, and *low* quality, based on the essays' grades. As a result, characteristics of essays belonging to specific groups should emerge, highlighting areas for specific improvement in low-quality essays, and features to be emulated in high-quality essays.

It was hypothesised that syntactic complexity could indicate general writing quality, as complex syntactic constructions can be used to convey complex ideas in writing. Furthermore, processing syntactically complex sentences places higher demands on working memory, particularly for students with a lower reading ability (Just and Carpenter, 1992). Lexical complexity was also hypothesised to indicate writing quality, serving as an indication of a student's range of vocabulary, and a general reflection of their linguistic ability (McNamara et al., 2010; McCarthy and Jarvis, 2007). The specific measurements of complexity used in this study are described in detail in section 2.2.

Similar studies have been carried out in this area in the past. For example, McNamara et al. (2010) examined the ability of various syntactic and lexical features to predict essay grades using a corpus containing 120 argumentative essays written by undergraduate university students, and Jagaiah (2017) analysed differences in syntactic complexity between argumentative essays written by at-risk and not-at-risk eighth-grade students using a dataset containing 1029 essays. Kumar and Boulanger (2020) uses explainable AI techniques to examine linguistic features of essay quality.

# 2 Method

## 2.1 Dataset

The Automated Student Assessment Prize (ASAP)* dataset contains collections of essays hand-scored by expert human graders. We chose to analyse the essays in *Essay Set #1*, containing 1782 essays. These essays are written in a persuasive, narrative, and expository style, responding to a prompt on the role of computers in society (see Appendix A for full prompt). The essays were written by eighth-grade students, as mentioned in the introduction. Essays were graded by two human graders according to a 6-point rubric, which is available in Appendix B. If both grade scores are adjacent, then the final score is the sum of those two scores. Otherwise, the final score is determined by an expert scorer. Thus, the grade for each essay is an integer in the interval $[2, 12]$. The grades have a negatively-skewed distribution, with $\mu = 8.53$ and $\sigma = 1.54$. Word count has a more symmetric distribution, with $\mu = 365.68$, $\sigma = 119.6$.

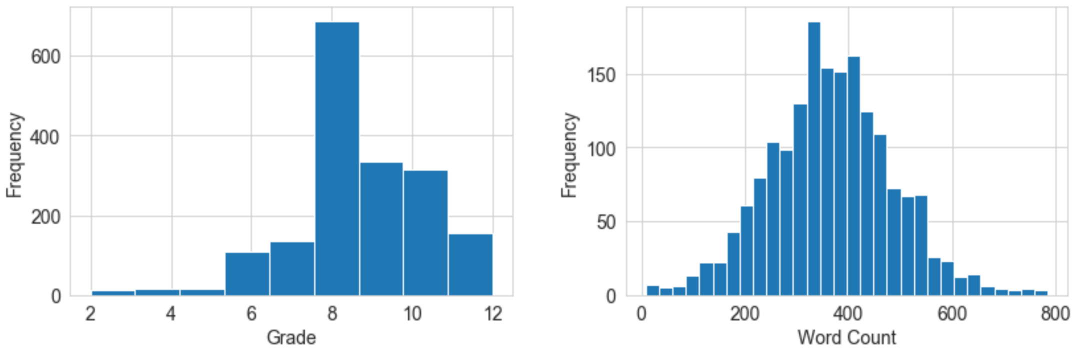


**Fig. 1.** Distributions for grade and word count of essays in *Essay Set #1*.

The essays were partitioned to form *proficiency groups*. The low-proficiency group contained essays with grades in the interval [2, 6], the medium-proficiency with grades in [7, 10], and the high-proficiency with grades in [11, 12]. This partition was designed to produce an approximately equal number of high-proficiency and low-proficiency essays, with each group containing 156 and 155 essays, respectively. A similar approach was taken by McNamara et al. (2010), who split their data into two groups, according to proficiency. This partition differs in that it aims to identify groups containing particularly strong, particularly weak, and middle-of-the-road essays, enabling a more granular account of variation in linguistic complexity.

*https://www.kaggle.com/c/asap-aes

## 2.2 Complexity Analysis

The following subsections describe the measures of syntactic and lexical complexity that were calculated for the essays in the dataset described above. Word count and sentence count were also calculated, and while neither are considered measures of syntactic or lexical complexity in this case (Davison and Kantor, 1982), they serve as simple, general measures of linguistic complexity.

### 2.2.1 Syntactic Complexity

A number of measures of syntactic complexity were used for this study. Namely, those measures included in the L2 Syntactic Complexity Analyzer (L2SCA) [†] (Lu, 2010), and the mean number of words before the main verb of the sentence, which was used as a measure of syntactic complexity by McNamara et al. (2010).

- **L2SCA:** this system computes fourteen measurements relating to structural syntactic complexity of a given piece of text. Developed by Lu (2010), the system consists of three steps: (i) analyzing the syntactic structure of the input text using the Stanford parser (Klein et al., 2003), (ii) using Tregex (Levy and Andrew, 2006) queries to count occurrences of relevant production units and syntactic structures, and (iii) calculating a number of measures of syntactic complexity based on those counts. The production units relevant to these measures described as follows, taken from Lu (2010):

  *Sentence*: a group of words followed by a full stop, question mark, exclamation mark, quotation mark, or ellipsis, indicating the end of the sentence (Hunt, 1965).

  *Clause*: a structure containing a subject and finite verb (Hunt, 1965).

  *Dependent clause*: a finite adjectival, adverbial, or nominal clause (Cooper, 1976; Hunt, 1965).

  *T-unit*: a main clause and any subordinate clause or non-clausal structure attached to or embedded in it (Hunt, 1970).

  *Complex T-unit*: a T-unit that contains a dependent clause (Casanave, 1994).

  *Coordinate phrase*: An adjective, adverb, noun, or verb phrase that immediately dominates a coordinating conjunction (Cooper, 1976).

  *Complex nominal*: either (i) noun plus adjective, possessive, prepositional phrase, relative clause, participle, or appositive, (ii) nominal clauses, or (iii) gerunds and infinitives in subject position (Cooper, 1976).

  *Verb phrase*: a finite or non-finite verb phrase.

[†] http://www.personal.psu.edu/xxl13/downloads/l2sca.html

Complexity measurements are typically a ratio of counts of two of the above described production units or syntactic structures. A summary of the measurements is given in Fig. 2.

| Measure | Definition |
|---|---|
| Mean length of clause | # words / # clauses |
| Mean length of sentence | # words / # sentences |
| Mean length of T-unit | # words / # T-units |
| Sentence complexity ratio | # clauses / # sentences |
| T-unit complexity ratio | # clauses / # T-units |
| Complex T-unit ratio | # complex T-units / # T-units |
| Dependent clause ratio | # dependent clauses / # clauses |
| Dependent clauses per T-unit | # dependent clauses / # T-units |
| Coordinate phrases per clause | # coordinate phrases / # clauses |
| Sentence coordination ratio | # T-units / # sentences |
| Complex nominals per clause | # complex nominals / # clauses |
| Complex nominals per T-units | # complex nominals / # T-units |
| Verb phrases per T-unit | # verb phrases / # T-units |

**Fig. 2.** Summary of L2SCA measures of syntactic complexity adapted from Lu (2010).

- **Words before verb:** another measurement of syntactic complexity used was the mean number of words appearing before the root verb of the sentence across an essay. This measure was found to be predictive of essay grade by McNamara et al. (2010). A python script was written to perform this calculation, using the Natural Language Toolkit (NLTK)[‡] (Bird et al., 2009) for sentence tokenization, and the Stanford CoreNLP Universal Dependencies parser (Schuster and Manning, 2016) to determine the root verb of each sentence. The root verb of a sentence is the root node of the Universal Dependencies parse tree. For example, see Fig. 3 for a visual representation of the dependency parse tree for the sentence *"Firstly, computers help teach hand eye coordination"*, taken from one of the essays in the dataset, in which the word *help* is the root verb.

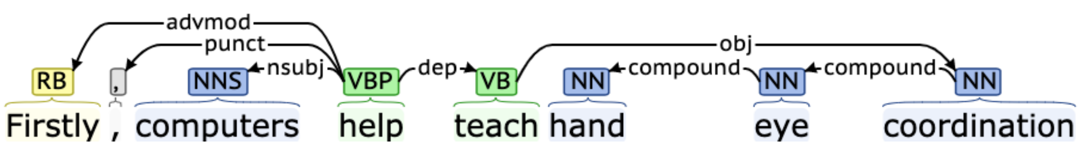


**Fig. 3.** Universal Dependencies parse tree for the sentence *"Firstly, computers help teach hand eye coordination"*. Visualisation was created using `corenlp.run`[§].

[‡]https://www.nltk.org/

### 2.2.2 Lexical Complexity

We use the three measures of lexical diversity that are recommended by McCarthy and Jarvis (2010): Measure of Textual Lexical Diversity (MTLD), Hypergeometric Distribution (HD-D), and Maas Type-Token Ratio (Maas TTR), as well as simple statistics on word length and syllable count. These measures were recommended as each captures unique lexical information. *Types* and *tokens* are terms relevant to lexical diversity which are essential to understanding MTLD and Maas TTR. The number of *tokens* in a piece of text is simply the word count. The number of *types*, on the other hand, is the number of unique words used in the text, i.e. the size of the set of words used. These figures can be combined to form the Type-Token Ratio (TTR) metric, $TTR = \frac{\#types}{\#tokens}$. TTR is not used directly in this study, however, as it is considered problematic due to its relationship with text length (McCarthy and Jarvis, 2010). The following contains descriptions of each measure of lexical complexity used in this study.

- **Measure of Textual Lexical Diversity (MTLD):** this measure involves calculating TTR for each word sequentially in the text until the TTR drops below a given threshold called the *factor size*. Once TTR is below the factor size, the *factor count* is increased by 1, and the TTR value is reset. The default factor size is 0.72. An example from McCarthy and Jarvis (2010):

  *of* (1.0) *the* (1.0) *people* (1.0), *by* (1.0) *the* (0.8) *people* (0.667) $||f := f+1||$, *for* (1.0), *the* (1.0) ...

  In this example, the TTR calculated at each word is presented in parentheses, and $f$ is factor count. *Partial factors* may also be added to the factor count when there is a remainder TTR that hasn't fallen below the factor size threshold. The proportion of the range between the factor size and 1.0 that this remainder occupies may be added to the factor count. Final MTLD is calculated as the mean of a forward pass MTLD and a backward pass MTLD.

- **Hypergeometric Distribution (HD-D):** *hypergeometric distribution* is a discrete probability distribution describing the probability of $k$ successful draws, without replacement, out of $n$ total draws from a population of size $N$ that contains $K$ objects with the feature that determines a successful draw. The probability mass function for this distribution is as follows (from Rice (2006)):

  $$P(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}} \tag{1}$$

  The hypergeomtric distribution can represent the probability of drawing $k$ tokens of a given type from a sample of $n$ tokens from the text. This is used in the calculation of the HD-D index, which involves computing, for each type in the text, the probability that one of more of its tokens would appear in a random sample of 42 tokens drawn from the text. The final value for HD-D is the sum of these probabilities for each type in the text (McCarthy and Jarvis, 2010). This is given by the following equation:

$$HDD = \sum_{t \in T} 1 - \frac{\binom{N-\#t}{42}}{\binom{N}{42}} \tag{2}$$

Where $T$ is the set of types, $\#t$ is the number of tokens of type $t$ in the text, and $N$ is the total number of tokens in the text.

- **Maas Type-Token Ratio (Maas TTR):** this metric attempts to minimise the impact of sample length on TTR using log functions as shown in the equation 3. (Fergadiotis et al., 2015; Maas, 1972). McCarthy and Jarvis (2010) showed the effect of sample length on Maas TTR to be just 1.5%.

$$.M = \frac{log(N) - log(\#T)}{log^2(N)} \tag{3}$$

  Using $N$ for text length and $T$ for set of types, as before.

- **Word Length:** the word length for each word in the text was calculated. The mean and maximum of the list of words lengths for each text were taken as measurements of lexical complexity. It is noteworthy that the maximum word length statistic may be impacted by word count: assuming a random selection of words, the addition of more words increases the likelihood of a large word being selected.

- **Syllable Count:** the number of syllables in each word was calculated, using the `syllables`[¶] python library. As with word length, the mean and maximum of these counts was calculated for each essay as a measure of lexical complexity.

# 3 Results

Each of the measurements described in the previous section was calculated for each essay in the dataset. The mean and standard deviation per proficiency group are presented, with proficiency groups labelled by 'high', 'medium', and 'low'. Additionally, a linear regression analysis was conducted for each complexity measure in order to examine its correlation with essay grade. This is summarised by the $R^2$ statistic, which indicates for a given measure of complexity the proportion of the variance in essay grades that is predictable from that measure. Violin plots and regression plots are used to visualise

[¶] https://pypi.org/project/syllables/

proficiency group distributions, differences between groups, and correlations between measurements and grades for those measures of complexity that show promising results.

## 3.1 General Complexity

General complexity measures included word and sentence counts. Frequency of syntactic constituents were also considered, although they would correlate highly with word count. Fig. 4 displays results for these general complexity measures.

| Measure | High | Medium | Low | $R^2$ |
|---|---|---|---|---|
| Word count | 529.17 (93.66) | 374.22 (96.27) | 167.52 (70.28) | 0.63 |
| Sentence count | 32.74 (6.7) | 23.14 (7.91) | 9.54 (5.75) | 0.48 |

**Fig. 4.** Results of calculations for general measures of complexity.

These results indicate that word count in particular is a good indicator of essay grade for this dataset, with longer essays typically being awarded higher grades. Word count results are visualised in Fig. 5.

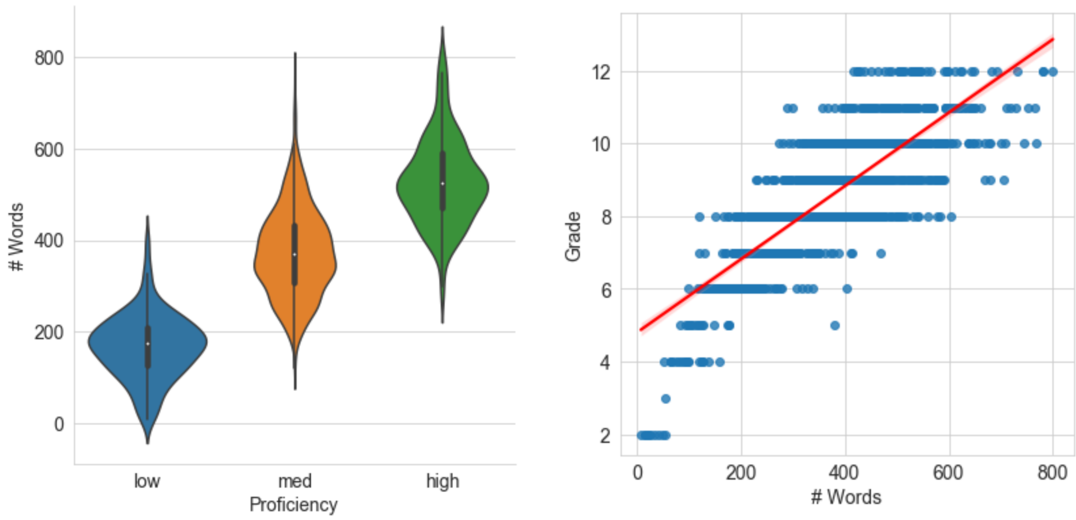


**Fig. 5.** Differences between groups and general correlation for word count measure of complexity.

## 3.2 Syntactic Complexity

Results for the various measures of syntactic complexity employed are presented in Fig. 6.

Differences in complexity among proficiency groups are evident in many of the measures presented above, although large differences in variance make some groups hard to compare accurately. For example, *mean length of sentence* has a standard deviation of 2.94 for high proficiency essays, but 13.45 for low proficiency essays. Promising measures are typically

| Measure | High | Medium | Low | $R^2$ |
|---|---|---|---|---|
| Mean words before verb | 4.72 (1.18) | 4.1 (1.4) | 3.76 (1.79) | 0.03 |
| Mean length of sentence | 16.5 (2.94) | 17.37 (5.74) | 21.3 (13.45) | 0.03 |
| Mean length of clause | 9.06 (1.25) | 8.36 (1.21) | 7.97 (1.7) | 0.06 |
| Mean length of T-unit | 15 (2.56) | 15.53 (4.57) | 18.04 (10.13) | 0.02 |
| Sentence complexity ratio | 1.84 (0.33) | 2.1 (0.72) | 2.7 (1.68) | 0.07 |
| T-unit complexity ratio | 1.67 (0.27) | 1.88 (0.58) | 2.29 (1.2) | 0.06 |
| Complex T-unit ratio | 0.46 (0.12) | 0.51 (0.16) | 0.58 (0.23) | 0.05 |
| Dependent clause ratio | 0.37 (0.08) | 0.41 (0.1) | 0.44 (0.15) | 0.04 |
| Dependent clause per T-unit | 0.63 (0.23) | 0.82 (0.5) | 1.12 (0.97) | 0.06 |
| Coordinate phrases per clause | 0.25 (0.1) | 0.23 (0.11) | 0.24 (0.19) | $< 0.01$ |
| Sentence coordination ratio | 1.1 (0.09) | 1.12 (0.13) | 1.19 (0.48) | 0.01 |
| Complex nominals per clause | 0.93 (0.22) | 0.82 (0.2) | 0.81 (0.25) | 0.03 |
| Complex nominals per T-unit | 1.55 (0.4) | 1.54 (0.6) | 1.83 (1.09) | 0.01 |
| Verb phrases per T-unit | 2.25 (0.38) | 2.52 (0.78) | 3.01 (1.66) | 0.05 |

**Fig. 6.** Results of calculations for measures of syntactic complexity.

those with higher $R^2$ values and similar standard deviations across proficiency groups, such as *mean words before verb*, *mean length of clause*, and *dependent clause ratio*. See Fig. 7 for visual representations of these measures.

### 3.3 Lexical Complexity

Results for the measures of lexical complexity used are given in Fig. 8.

With the exception of *Maas TTR*, lexical complexity appears to behave somewhat uniformly across groups for each measure, with higher quality essays showing higher lexical complexity, and lower quality essays showing lower lexical complexity. This trend is depicted in Fig. 9, in which the values for each complexity measurement are normalised using min-max normalization for easier comparison. The strongest indicator appears to be *HD-D*. *MTLD* and *max number of syllables* also appear to have some predictive power. Visual representations of these measurements are given in Fig. 10. Interestingly, max word length and number of syllables show much greater predictive power than mean word length and number of syllables.

## 4 Discussion

An issue mentioned in section 3.2 is that many measures of complexity show a large difference in standard deviation across proficiency groups. This difference is caused by outliers, which have the effect of pulling the mean up, and causing an explosion in standard deviation. One such outlier is an essay in which no punctuation was used, resulting in the entire essay being recognised as a single sentence. The following is an excerpt from that essay.

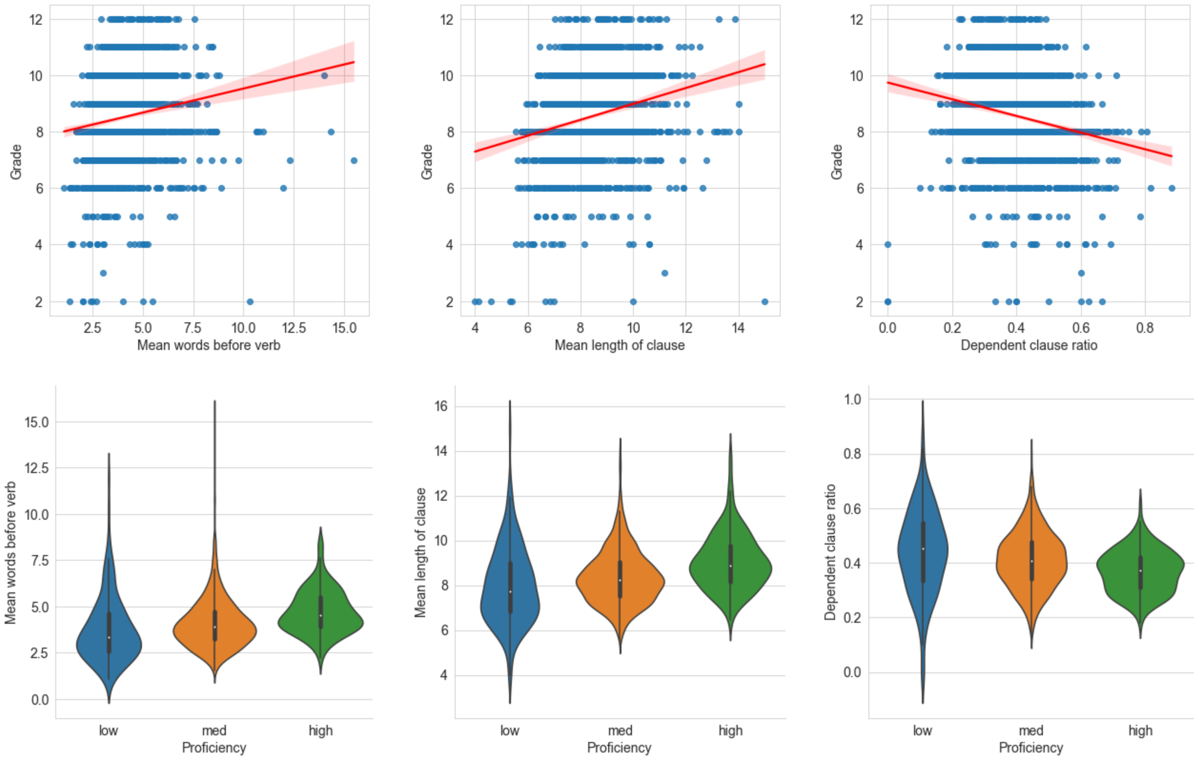


**Fig. 7.** Differences between groups and correlation with grade for *mean words before verb*, *mean length of clause*, and *dependent clause ratio*.

| **Measure** | **High** | **Medium** | **Low** | $R^2$ |
|---|---|---|---|---|
| Maas TTR | 0.05 (0.005) | 0.06 (0.008) | 0.06 (0.01) | 0.06 |
| HD-D | 0.83 (0.02) | 0.8 (0.03) | 0.77 (0.05) | 0.23 |
| MTLD | 78.59 (15.6) | 65.03 (14.86) | 54.82 (18.09) | 0.16 |
| Mean word length | 4.09 (0.2) | 3.98 (0.2) | 3.95 (0.26) | 0.03 |
| Max word length | 13.5 (1.68) | 12.55 (1.85) | 10.99 (1.59) | 0.12 |
| Mean number of syllables | 1.47 (0.08) | 1.42 (0.08) | 1.4 (0.1) | 0.04 |
| Max number of syllables | 5 (0.61) | 4.62 (0.65) | 4.05 (0.63) | 0.15 |

**Fig. 8.** Results of calculations for measures of lexical complexity.

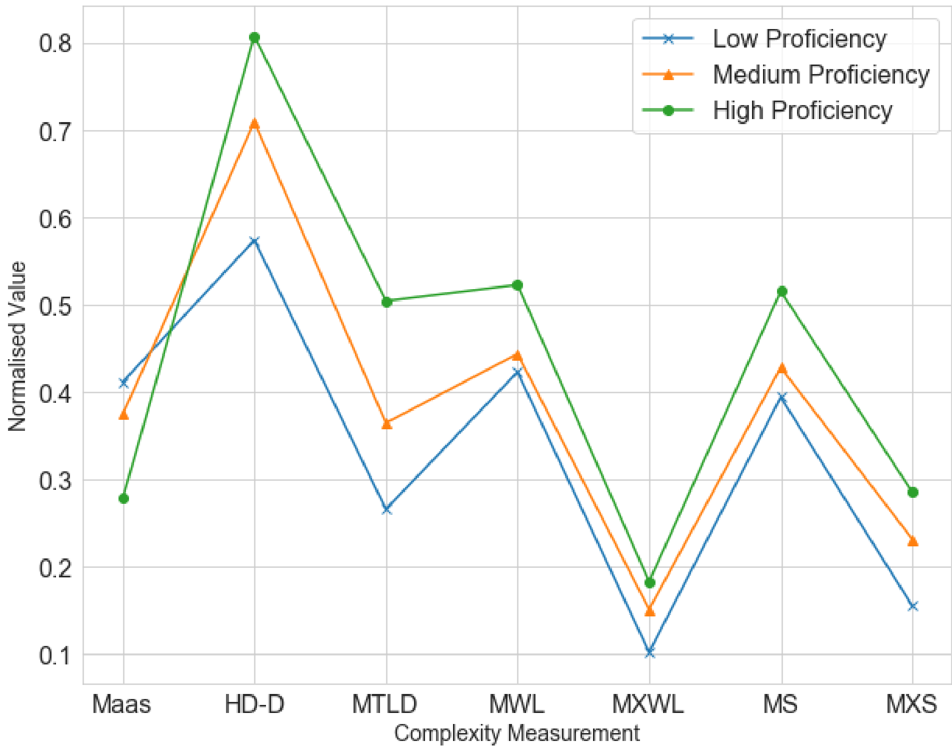


**Fig. 9.** Mean lexical complexity is shown to be higher in high-quality essays, and lower in low-quality essays, for most measures.

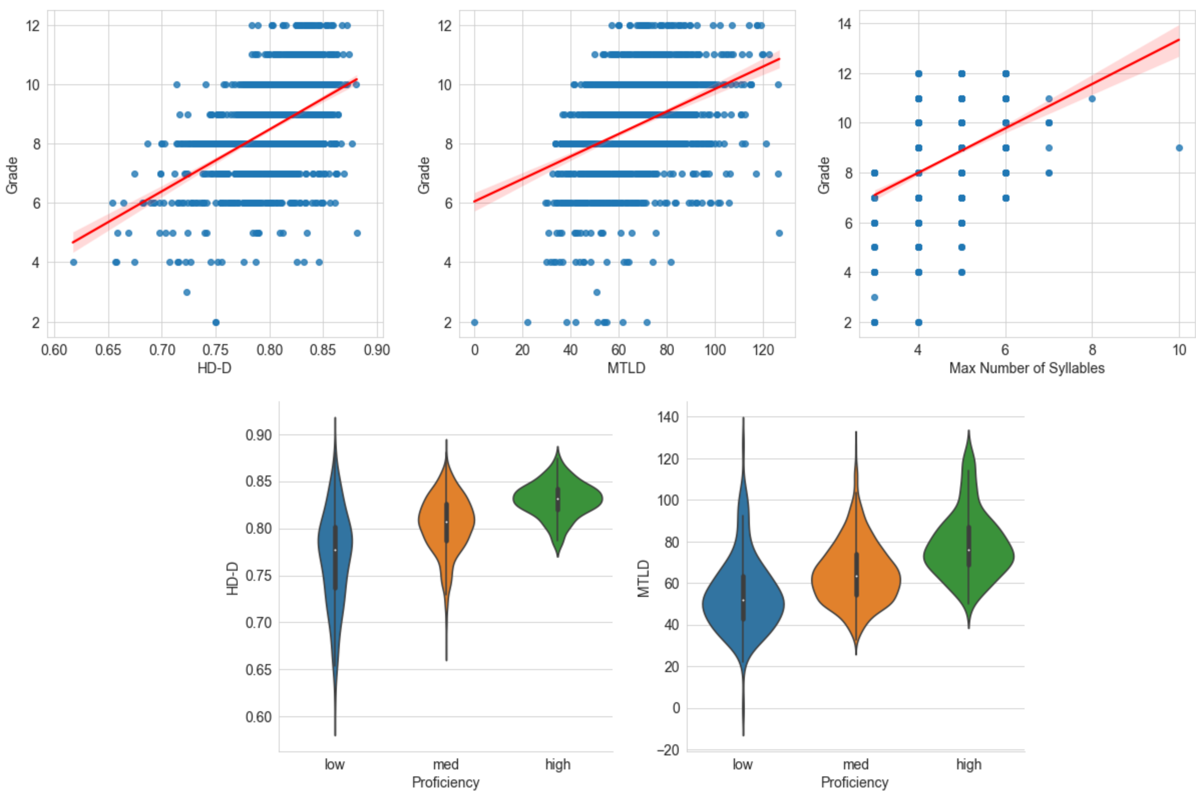


**Fig. 10.** Differences between groups for *HD-D* and *MTLD*, and regression plots for *HD-D*, *MTLD*, and *max number of syllables*.

> ”Computers don’t have any affect on kids we just love going on cause we use it for help and this persuade the readers of the local newspaper cause we need to be able to communicate also do writing essays and doing social studies or science homework my ideas are let us go computers cause were not bothering u can just leave us alone ...”

This exposes a weakness in *mean length of sentence* as a measure of complexity for poor quality writing, namely that it relies on correct use of punctuation. This weakness can be seen throughout the per-sentence and per-T-unit measurements, each of which inherit a reliance on correct use of punctuation. In fact, these outliers caused the apparent correlation to reverse in some cases, such that lower quality essays had higher levels of syntactic complexity. A more resilient measure is *mean length of clause*, for which each proficiency group shows more similar standard deviations. These results suggest that, in general, if comparing texts that span a wide range of writing proficiency, *mean clause length* is preferable to *mean sentence length*. Another valuable measure is *mean words before verb*, which in this experiment yielded similar results to those of McNamara et al. (2010). *Dependent clause ratio* shows a negative correlation with essay grade, indicating that clauses in high-quality essays are more likely to express complete thoughts[‖].

Overall, measurements of lexical complexity show stronger predictive power than those of syntactic complexity. Distributions are much more similar across groups, and $R^2$ values are higher almost across the board. *Max number of syllables* was one of the more powerful predictors, and interestingly showed better results than *max word length*, despite their similarity. It is evident from the regression plot for max number of syllables in Fig. 10, not a single essay in the low-proficiency group contained a word with more than 5 syllables, while no essays in the high-proficiency group had a max syllable word with less than 4 syllables. It is noteworthy that *max word length* and *max number of syllables* may be affected by essay length, and so a measure such as *mean of max syllables per sentence* may be more indicative of general lexical complexity.

Measures of linguistic complexity that are found to correlate with essay writing quality, such as those presented above, may be used to generate personalised pedagogical feedback. For example, a student whose essay has low measures of lexical complexity might be encouraged to read more in order to build a wider vocabulary (Cain and Oakhill, 2011). Similarly, these measures could be used to create writing strategies according to proficiency level (De Silva and Graham, 2015).

[‖] https://owl.purdue.edu/owl/general_writing/punctuation/independent_and_dependent_clauses/

# 5 Conclusion

This study indicates that various measures of linguistic complexity can predict writing quality in persuasive essays written by eighth-grade students. In general, higher linguistic complexity indicated higher writing quality. The strongest indication of quality was, in this case, essay word count. Measures of lexical complexity, and in particular *HD-D*, also served as powerful predictors. Measures of syntactic complexity showed some predictive power, although less than those of lexical complexity. Additionally, many measures of syntactic complexity were negatively affected by improper use of punctuation in the low-proficiency groups. These findings can be incorporated in areas such as pedagogical feedback and automated essay scoring. Python notebooks containing complexity analysis and correlation evaluations are available on github[**].

Future research in this area might look towards measuring changes in complexity throughout written essays, potentially identifying specific sections of the text that negatively impacted the overall writing quality.

[**] https://github.com/OisinNolan/Writing-Quality-Prediction

# References


Bird, S., Klein, E., and Loper, E. (2009). *Natural language processing with Python: analyzing text with the natural language toolkit.* " O'Reilly Media, Inc.".

Cain, K. and Oakhill, J. (2011). Matthew effects in young readers: Reading comprehension and reading experience aid vocabulary development. *Journal of learning disabilities*, 44(5):431–443.

Casanave, C. P. (1994). Language development in students' journals. *Journal of second language writing*, 3(3):179–201.

Cooper, T. C. (1976). Measuring written syntactic patterns of second language learners of german. *The Journal of Educational Research*, 69(5):176–183.

Davison, A. and Kantor, R. N. (1982). On the failure of readability formulas to define readable texts: A case study from adaptations. *Reading research quarterly*, pages 187–209.

De Silva, R. and Graham, S. (2015). The effects of strategy instruction on writing strategy use for students of different proficiency levels. *System*, 53:47–59.

Fergadiotis, G., Wright, H. H., and Green, S. B. (2015). Psychometric evaluation of lexical diversity indices: Assessing length effects. *Journal of Speech, Language, and Hearing Research*, 58(3):840–852.

Ferretti, R. P. and Graham, S. (2019). Argumentative writing: Theory, assessment, and instruction. *Reading and Writing*, 32(6):1345–1357.

Hunt, K. W. (1965). Grammatical structures written at three grade levels. ncte research report no. 3.

Hunt, K. W. (1970). Do sentences in the second language grow like those in the first? *Tesol Quarterly*, pages 195–202.

Jagaiah, T. (2017). Analysis of syntactic complexity and its relationship to writing quality in argumentative essays.

Just, M. A. and Carpenter, P. A. (1992). A capacity theory of comprehension: individual differences in working memory. *Psychological review*, 99(1):122.

Klein, D., Manning, C. D., et al. (2003). Fast exact inference with a factored model for natural language parsing. *Advances in neural information processing systems*, pages 3–10.

Kumar, V. and Boulanger, D. (2020). Explainable automated essay scoring: Deep learning really has pedagogical value. In *Frontiers in Education*, volume 5, page 186. Frontiers.

Levy, R. and Andrew, G. (2006). Tregex and tsurgeon: tools for querying and manipulating tree data structures. In *LREC*, pages 2231–2234. Citeseer.

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International journal of corpus linguistics*, 15(4):474–496.

Maas, H.-D. (1972). Über den zusammenhang zwischen wortschatzumfang und länge eines textes. *Zeitschrift für Literaturwissenschaft und Linguistik*, 2(8):73.

McCarthy, P. M. and Jarvis, S. (2007). vocd: A theoretical and empirical evaluation. *Language Testing*, 24(4):459–488.

McCarthy, P. M. and Jarvis, S. (2010). Mtld, vocd-d, and hd-d: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior research methods*, 42(2):381–392.

McNamara, D. S., Crossley, S. A., and McCarthy, P. M. (2010). Linguistic features of writing quality. *Written communication*, 27(1):57–86.

McNamara, D. S., Graesser, A. C., McCarthy, P. M., and Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press.

Rice, J. A. (2006). *Mathematical statistics and data analysis*. Nelson Education.

Schuster, S. and Manning, C. D. (2016). Enhanced english universal dependencies: An improved representation for natural language understanding tasks. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 2371–2378.

# A Prompt for Essay Set #1

"More and more people use computers, but not everyone agrees that this benefits society. Those who support advances in technology believe that computers have a positive effect on people. They teach hand-eye coordination, give people the ability to learn about faraway places and people, and even allow people to talk online with other people. Others have different ideas. Some experts are concerned that people are spending too much time on their computers and less time exercising, enjoying nature, and interacting with family and friends.

Write a letter to your local newspaper in which you state your opinion on the effects computers have on people. Persuade the readers to agree with you."

# B Essay Grading Rubric

**Score Point 1:** An undeveloped response that may take a position but offers no more than very minimal support. Typical elements:

- Contains few or vague details.
- Is awkward and fragmented.
- May be difficult to read and understand.
- May show no awareness of audience.

**Score Point 2:** An under-developed response that may or may not take a position. Typical elements:

- Contains only general reasons with unelaborated and/or list-like details.
- Shows little or no evidence of organization.
- May be awkward and confused or simplistic.
- May show little awareness of audience.

**Score Point 3:** A minimally-developed response that may take a position, but with inadequate support and details. Typical elements:

- Has reasons with minimal elaboration and more general than specific details.
- Shows some organization.
- May be awkward in parts with few transitions.
- Shows some awareness of audience.

**Score Point 4:** A somewhat-developed response that takes a position and provides adequate support. Typical elements:

- Has adequately elaborated reasons with a mix of general and specific details.
- Shows satisfactory organization.

- May be somewhat fluent with some transitional language.
- Shows adequate awareness of audience.

**Score Point 5:** A developed response that takes a clear position and provides reasonably persuasive support. Typical elements:

- Has moderately well elaborated reasons with mostly specific details.
- Exhibits generally strong organization.
- May be moderately fluent with transitional language throughout.
- May show a consistent awareness of audience.

**Score Point 6:** A well-developed response that takes a clear and thoughtful position and provides persuasive support. Typical elements:

- Has fully elaborated reasons with specific details.
- Exhibits strong organization.
- Is fluent and uses sophisticated transitional language.
- May show a heightened awareness of audience.