

## PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<https://hdl.handle.net/2066/226349>

Please be advised that this information was generated on 2021-11-05 and may be subject to change.



# Syntactic predictors for text quality in Dutch upper-secondary school students' L1 argumentative writing

Jimmy H. M. van Rijt<sup>1,2</sup> · Brenda van den Broek<sup>3</sup> · Sven De Maeyer<sup>4</sup>

© The Author(s) 2020

## Abstract

Among other things, learning to write entails learning how to use complex sentences effectively in discourse. Some research has therefore focused on relating measures of syntactic complexity to text quality. Apart from the fact that the existing research on this topic appears inconclusive, most of it has been conducted in English L1 contexts. This is potentially problematic, since relevant syntactic indices may not be the same across languages. The current study is the first to explore which syntactic features predict text quality in Dutch secondary school students' argumentative writing. In order to do so, the quality of 125 argumentative essays written by students was rated and the syntactic features of the texts were analyzed. A multilevel regression analysis was then used to investigate which features contribute to text quality. The resulting model (explaining 14.5% of the variance in text quality) shows that the relative number of finite clauses and the ratio between the number of relative clauses and the number of finite clauses positively predict text quality. Discrepancies between our findings and those of previous studies indicate that the relations between syntactic features and text quality may vary based on factors such as language and genre. Additional (cross-linguistic) research is needed to gain a more complete understanding of the relationships between syntactic constructions and text quality and the potential moderating role of language and genre.

**Keywords** Argumentative writing · Syntactic complexity · Text quality · Secondary education · Dutch language

---

✉ Jimmy H. M. van Rijt  
j.vanrijt@fontys.nl; j.vanrijt@let.ru.nl

Extended author information available on the last page of the article

## Introduction

The importance of students' writing skills is not easily overestimated. Writing plays a major role in modern society on so many levels, that Graham and Perin (2007, p. 445) conclude that 'adolescents who do not learn to write well are at a disadvantage' and that 'in school, weaker writers are less likely than their more skilled classmates to use writing to support and extend learning in content classrooms.' This means that teaching students how to write well should be an essential part of language education, especially because there are many concerns about the current level of students' writing skills (Graham & Perin, 2007; MacArthur, Graham & Fitzgerald, 2016).

As the influential models of Hayes and Flower (1980) and Bereiter and Scardamalia (1987) have shown, the process of writing calls upon various types of skills and knowledge, each of which important in its own right. One of these types of knowledge is linguistic knowledge, which includes the role of grammar in texts. One question that emerges when thinking about the role of grammar in writing, is whether there are relationships between specific syntactic features and text quality. If such relations exist, they could inform writing education. In this article, grammatical features mainly pertain to syntax. Grammatical and syntactic features should thus be seen as synonyms throughout.

Previous studies on the relation between grammatical features of students' L1 writing and the quality of their texts have produced mixed results. While several studies report links between essay quality and syntactic complexity (e.g., Myhill, 2008), other studies find no significant relationships (e.g., McNamara et al., 2013; Perin & Lauterbach, 2016). In addition, the majority of the existing research on the relationship between grammatical features of a text and its quality has either focused on primary education (e.g., Benson & Campbell, 2009; Klecan-Aker & Hendrick, 1985) or on higher education (e.g., McNamara, Scottley & McCarthy, 2010; Perin & Lauterbach, 2016), leaving secondary school contexts understudied (Myhill, 2008). Finally, very little is known about the potentially moderating role of any specific language, since most research on this topic has been carried out in contexts in which English is the students' L1 (with the exception of the Spenser project, cf. next section). Much research is therefore still needed to gain an adequate picture of the relationships between syntactic features of a text and text quality (cf. Crowhurst, 1980; MacArthur et al., 2019). The current study aims to contribute to the existing knowledge on this topic by examining, in a Dutch secondary school context, the relation between text quality and syntactic features that have been investigated in other educational jurisdictions with a different L1.

### Syntactic complexity and writing quality

As previously mentioned, the idea that there are links between students' syntactic development or competence and their ability to construct strong texts is not novel. The majority of studies on the relationship between syntactic constructions and writing has focused on the concept of syntactic complexity (e.g., Chall & Dale, 1995; Kleijn, 2018; McNamara, Scottley & McCarthy, 2010; Pitler & Nenkova, 2008).

This concept is an interesting one since on the one hand it is associated with slow processing and impaired understanding (Gibson, 1998), while on the other hand it is seen as a characteristic of sophisticated language use (Kyle, 2016; McNamara et al., 2010). This paradox explains why studies relating syntactic complexity to readability tend to find rather different results when compared to studies relating syntactic complexity to text quality.

Generally, readability studies find that measures of syntactic complexity are negatively related to the readability of a text (e.g., Chall & Dale, 1995; Crossley, Greenfield & McNamara, 2008). In other words: texts that are more syntactically complex are less readable. Studies on text quality, however, tend to find that syntactically more complex texts are rated to be of higher quality (e.g., Crossley, Weston, McLain Sullivan & McNamara, 2011; McNamara et al., 2010; Myhill, 2008; 2009). Myhill (2008), for instance, examined 718 pieces of writing of secondary school students (aged 13–15) and found that as student age, they are able to use more complex and advanced constructions. In addition, more complex language use (such as using variation in sentence length and sentence patterning) was found to be an important characteristic of strong texts. Furthermore, in what has become known as ‘the Spenser project’, relations between complex language use and writing ability were also found. The findings from this project show that, across languages, more advanced students tend to use more complex sentences (Berman & Verhoeven, 2002) and that adults tend to show an increased competence in formulating complex sentences compared to younger writers (Verhoeven et al., 2002).

Texts in which more syntactically complex sentence structures are used thus seem to be perceived as more advanced, even though they might be less readable. However, a recent study by MacArthur et al. (2019) pointed towards a negative relationship between syntactic complexity and text quality. The authors state that this finding is surprising and contradicts previous work. Based on, amongst others things, this observation, MacArthur et al. (2019) state that ‘much further research is needed to confirm and extend the present findings.’

## Measures of syntactic complexity

Throughout the existing studies, syntactic complexity is measured in varying ways. According to MacArthur et al. (2019), one of the challenges of using linguistic analysis to predict text quality is the enormous number of available linguistic indices. In existing research, numerous different indices have been investigated in relation to writing quality. First of all, many readability formulas, such as the ones by Chall and Dale (1995) and Staphorsius (1994), measure syntactic complexity by means of sentence length (Kleijn, 2018). The advantage of this measure is that it is a non-ambiguous measure which can easily and reliably be established (Kyle, 2016). However, it is also a rather unintelligent measure which some have deemed too shallow to actually capture the concept of syntactic complexity (Kleijn, 2018). Therefore, other scholars have used more complex measures. McNamara et al. (2010), for instance, used the number of words before the main verb as an indicator of syntactic complexity and found it to be positively related to text quality of undergraduate students.

Another measure that has been used is the use of the passive voice (MacArthur et al., 2019; Renkema, 2004). Kameen (1979) used this measure and found that texts of higher quality generally contain more passive constructions. This finding concurs with the findings of Myhill (2008) who found that strong writers tend to use more passive constructions than weak writers. Meanwhile, many writing handbooks recommend avoiding the use of passive constructions for the sake of readability (for Dutch, see Renkema, 2012, p. 92). Indeed, increased passive voice use threatens a text's readability (Ferreira, 2003).

Another measure used in previous work is the complexity of noun phrases (NP's), which has been labeled 'the number of modifiers per noun phrase' or 'length of nominal phrases' by some (e.g., Crossley et al., 2011; MacArthur et al., 2019). The idea behind the measure is that the more a core noun in an NP is modified, the more it is an indication of complex language use. For example, the NP 'voting behaviour' is a compound, consisting of two nouns. Such an NP can easily be modified to increase its syntactic complexity, for example by modifying it with an adjective: 'unwanted voting behaviour'. The NP can be modified further when a prepositional phrase (PP) is added: 'unwanted voting behaviour of students'. According to Biber, Gray and Poonpon (2011), complex noun phrases are 'the hallmark of academic writing', and therefore of great importance. And indeed, several studies find that more complex NP's correlate with more advanced writing (e.g., Crossley et al., 2011; Haswell, 1990; MacArthur et al., 2019) and, in concordance with that, Myhill (2008) found that moderate writers tend to use more 'adjectives and adverbs for explanation' than weak writers, which could also be seen as an indication of more complicated NP's, since NP's can be readily modified by such means.

A final measure related to syntactic complexity deals with subordination, and the relative amount of dependent clauses in sentences (cf. Hudson, 2009; Kyle, 2016). According to Renkema (2004, p. 153) and Kyle (2016, p. 18), measures related to subordinate clauses are often used in stylistic and text quality research. Since subordinate sentences involve imbedding one structure into another (recursion), such sentences can be deemed more complex than simple sentences. They are also considered more advanced than coordinating constructions, in which two or multiple main clauses are linked together using coordinate conjunctions, such as 'and'. Verhoeven et al. (2002) found that children tend to use more coordinate constructions than adults. They explain this finding by assuming that linking autonomous clauses is conceptually simpler than the linking of embedded clauses. This finding aligns with the results of Myhill (2008), who showed that weaker writers tend to prefer coordination over subordination. In particular, they seem to chain ideas together with coordinate conjunctions, particularly 'and', showing a limited command of sentence boundaries and within-sentence connectivity. Verhoeven et al. (2002) also found that adults use more adverbial, complement and relative clauses in their writing. Dependent clauses (in their various forms) may therefore be seen as a characteristic of good writing.

Different types of dependent clauses can have different effects on the quality of the text. Dependent clauses can be either finite or non-finite (infinitive clauses), and they can be 'normal' dependent sentences or they can be relative clauses. Myhill (2008) found that better writing seems to be associated with lower use of finite verbs

compared to weaker writing. In her study, it was found that ‘sentence structure in good writing is syntactically elaborated beyond simple subject-verb patterns, for example, by greater use of adverbs or non-finite clauses, or expanded noun phrases’ (p. 280). Hence, dependent non-finite clauses seem to be favored in texts over finite clauses. This finding corresponds with Haswell’s (1986), who reported that graduate students used more infinitives than undergraduate students. Finally, relative clauses, which are used to modify nouns by adding extra information to them in sentences which are commonly introduced by relative pronouns, are considered more mature writing (Verhoeven et al., 2002). This may be because according to Myhill (2008), more mature writers develop ‘a greater facility for expansion of ideas within sentences’, and relative clauses are used to do just that. Additionally, relative clauses are generally difficult to process (Mak, 2001), which could account for the construction’s association with a certain maturity.

### The Dutch context

Given the linguistic differences between languages, the findings from the English context cannot simply be assumed to hold for other language as well. Within the context of Dutch writing education, very limited research has been conducted on the relationships between syntactic features and writing quality. For primary education, the only research investigating students’ progression of grammatical complexity is the study by Van Til, Van Weerden, Hempken & Keune (2014). In their study, these authors compared texts of 9-year-olds to texts of 12-year-olds and found that the two groups produced syntactic units of equal length but that 12-year-olds used roughly twice as many imbedded clauses as the 9-year-olds. In addition, 12-year-olds made less use of the coordinate conjunction ‘en’ (‘and’), which may indicate that students develop a more varied repertoire of conjunctions over time. For secondary education, there is (to the best of our knowledge) no work that explores relationships between syntactic constructions and writing ability. Insight in these relations is, however, important because such understanding can be used to inform writing education.

Since Dutch language teachers spend limited amounts of time on writing education due to overloaded teaching schedules (Henkens, 2010), it would be helpful to establish ‘syntactic priorities’ they can use when providing students with feedback on their writing. According to a study in the context of Dutch secondary education by Ekens & Meestringa (2013), 14% of teachers’ comments on students’ writing relates to sentence build-up and syntactic constructions. However, formal documents describing how students should progress in terms of their language skills (Ekens & Meestringa, 2013; Meijerink et al., 2009) fail to clearly describe which aspects of sentence construction should be mastered at a certain level (Hoogeveen, 2017; Van Silfhout, 2018). This means that Dutch teachers do not possess a framework that they can rely on when commenting on students’ syntactic choices in writing or for effectively teaching syntactic choices in writing. Similarly, for students it often remains unclear how they might develop their ability to craft sentences which are not only grammatically correct, but also befitting

of the genre and the intended goal of the text they are writing. At the same time, students are expected to show increased mastery of (complex) sentences as they progress through education.

## Current study

In summary, no research to date has investigated the relationship between syntactic features and writing quality in the Dutch secondary school context, and official curriculum guidelines fail to provide teachers with more details about relevant syntactic features for writing education. The current study is the first to explore whether constructions found to be of importance for writing quality in English are of similar importance in Dutch, while also taking into account other potentially relevant syntactic indices (cf. Table 1). While there are many other potentially interesting linguistic indices, the current study thus focuses on syntactic features.

**Table 1** Measures of syntactic complexity

<i>Sentence characteristics</i>	<i>Operationalization</i>
Sentence length	Average sentence length in words
Variation in sentence length	Standard deviation of sentence length in words
<i>Subordinate clauses</i>	
Relative number of finite subordinate clauses	Number of finite subordinate clauses/number of words
Relative number of infinitive subordinate clauses	Number of infinitive subordinate clauses/number of words
Relative number of Relative clauses (RRC's)	Number of relative clauses (from relative pronoun till end of RC-sentence)/number of words
<i>Clause ratio</i>	
RF-ratio	Number of relative clauses/number of finite subordinate clauses
IF-ratio	Number of infinitive clauses/number of finite subordinate clauses
<i>Noun phrase complexity</i>	
Length of noun phrases (NP's)	Average noun phrase length in words* (only taking into account non-embedded NP's)
<i>Passive voice</i>	
Relative number of passives	Number of passives (including semi-passives)/number of words
<i>Words before main verb</i>	
Number of words before main verb	Average number of words before the main verb (excluding coordinate conjunctions)

\*Some theoretical linguists distinguish between NP's (Noun Phrases) and DP's (Determiner Phrases) (see Broekhuis & Keizer, 2012). However, for the current paper, this distinction was considered irrelevant. Therefore, determiners were taken into account when calculating the average NP length (e.g., *De ouders* ('the parents') would be considered a noun phrase consisting of two words)

## Method

### Participants

A total of 136 tenth grade students participated in this study (67 male, average age 15.5). They were registered at five different Dutch secondary schools. One class of each school participated. Dutch (or a Dutch dialect) was the first language of 130 of those students. Ten of the participating students had a language disorder such as dyslexia. Classes were selected for participation by their teachers.

### Materials

An argumentative essay writing task was developed for this study. In Dutch secondary education, writing an argumentative text is compulsory in the upper levels of education, which means that this text genre is an obvious candidate for the current investigation. For the task, students could choose to argue in favour of or against one of two statements: (A) Parents should have complete access to their children's internet behavior, or (B) The voting age should be lowered from 18 to 16. These statements were selected from the database of the Dutch debating institute ([www.debatinstituut.nl](http://www.debatinstituut.nl)) which includes validated debating statements. Both of the selected statements have a difficulty level of 2 out of 4, as established by the debating institute. Students were instructed to write 400–500 words and to include at least two arguments in their essay. Using secondary sources was not allowed. Multiple teachers confirmed that this writing task was appropriate for tenth grade students.

### Procedure

The participating students carried out the essay writing task on computers at their own secondary school. Before the actual writing task, a short online demographics questionnaire was completed by all participants. All students received identical textual instruction on the writing task and had 50 min to finish their essays. They were specifically instructed to carry out the task as well as they could within this time frame and to hand in a complete text.

### Text quality rating

The text quality of 133 texts was rated. One of the texts was not handed in correctly and proved untraceable; two of the texts were excluded because they were considered too short (i.e., containing less than 200 words). The quality of the texts was rated by means of D-PAC: an online platform for comparative judgement (<http://www.d-pac.be/>). In comparative judgement, raters repeatedly compare two performances (in our case texts) and decide which of the two is better. Performances are compared multiple times to various other performances by



multiple raters. Finally, this results in a scale that ranks all the rated performances from worst to best (Lesterhuis, Verhavert, Coertjens, Donche & De Maeyer, 2016). Rating the quality of the texts this way has several important advantages over other methods of text quality rating such as the use of rubrics (as used by, for instance, MacArthur et al., 2019 and McNamara et al., 2010; Pollit, 2012). First, this method encourages raters to rate the quality of a text as a whole (i.e., holistically) instead of using a finite set of criteria relating to certain aspects of the text (i.e., analytically). The latter has been criticized for being too systematic and constraining, thereby endangering the validity of the judgement (Sadler, 2009). Lesterhuis et al. (2018) show that when argumentative texts are rated by means of comparative judgement, raters consider a wide range of relevant text quality aspects. They conclude from this that comparative judgement is a valid way to assess the quality of (argumentative) texts. Besides the advantage regarding validity, comparative judgement also eliminates complications resulting from sequential effects and differences in the severity of raters (Pollitt, 2012).

The texts in the current study were rated by a group of 11 raters consisting of teacher trainers, teachers of Dutch, and researchers in the field of linguistics or discourse studies. They had an average of 7 years of experience in the assessment of texts. On average, texts were compared 24.3 times, resulting in a reliability coefficient of 0.86 (see Verhavert, De Maeyer, Donche and Coertjens, 2017).

### Analysis of syntactic measures

Table 1 provides an overview of the measures of syntactic complexity taken into account in the current study. The measures fall into six categories: sentence characteristics, subordinate clauses, clause ratio, noun phrase complexity, passive voice, and words before the main verb. The sentence characteristics were analyzed using the online tool Analyze My Writing (<http://www.analyzemywriting.com/>). The other predictors were analyzed manually by teachers of Dutch in expert teams which were trained to analyze the texts on certain aspects. The analyses were checked (and, where necessary, corrected) twice: once by another team member and once by the authors. For the statistical analyses, relative measures were calculated (see Table 1 for details), which are independent of text length. This is essential since texts length tends to be a strong predictor of text quality and analyses of other measures therefore need to control for this relation (MacArthur et al., 2019). In order to find out to what extent the relationships between different types of clauses influenced text quality, we also calculated relative measures on clause level: one variable concerned the ratio between finite clauses (the 'default' type of clause) and relative clauses (the so called RF-ratio). Another variable concerned the ratio between finite clauses and infinitive clauses (dubbed the IF-ratio). Syntactic features were taken into account regardless of whether they violated the linguistic norms. For example, a relative clause in Dutch can start with various relative pronouns, such as *dat*, *wat* or *die*, but even if students had chosen the wrong pronoun, or if any other violations of the linguistic norms were present in their writing, we ignored these aspects in the analyses.

## Statistical analyses

Statistical analyses were applied to 125 texts; 8 of the 133 rated texts were excluded from the analyses since they did not clearly argue in favour or against one of the two statements. First descriptive statistics were explored, correlations between syntactic measures were investigated, and possible effects of chosen statement (internet behavior or voting age) and position (in favour or against) on the syntactic measures were examined. Next, a linear regression analysis was carried out with text quality as the outcome measure and the syntactic measures as predictors. Since the students in our study were clustered in five different schools, the approach proposed by Sommet and Morselli (2017) was used to assess the proportion of variability in text quality that lies between schools. The intraclass correlation coefficient (ICC) was 0.10. This indicates that between-school differences explain 10% of text quality. Since even small ICC's are considered reason for multilevel modeling (e.g., Nezlek, 2008), we used a multilevel approach in our regression analysis. In our model, we assumed a random effect of school on intercept (i.e., we assumed that some schools generally contain stronger students than others) and a fixed slope (i.e., we assumed that the effect of the syntactic predictors on text quality is the same for students from different schools).

## Results

### Exploration of the data

Table 2 shows how often students chose to argue for or against statements A (Parents should have complete access to their children's internet behavior) and B (The voting age should be lowered from 18 to 16). The mean quality of the texts did not differ between statements ( $t(123)=0.08$ ,  $p=0.94$ ) or between positions ( $t(123)=0.69$ ,  $p=0.50$ ). On average, students wrote texts of 420 words ( $SD=79$ ). The number of words students wrote did not differ based on the statement they chose ( $t(123)=1.17$ ,  $p=0.25$ ) or the position (i.e., in favour of or against) they took ( $t(123)=0.03$ ,  $p=0.98$ ).

Table 3 provides an overview of the means and standard deviations of the syntactic measures investigated in the current study. There were significant differences regarding Noun phrase length ( $t(123)=3.87$ ,  $p<0.001$ ), the Relative number of passives ( $t(123)=3.46$ ,  $p<0.001$ ) and RF-ratio ( $t(123)=2.35$ ,  $p<0.05$ ) between texts on the two different statements. There were also significant differences regarding

**Table 2** Distribution of participants over statements and positions

Statement	N	In favour of	Against
A (internet behavior)	72	11	61
B (voting age)	53	28	25
Total	125	39	86

**Table 3** Means (standard deviations) of syntactic measures

	Absolute average per text	/Words	/Finite clauses
<i>Sentence characteristics</i>			
Sentence length	20.33 (5.96)	n.a.	n.a.
Variation in sentence length	9.85 (3.97)	n.a.	n.a.
<i>Subordinate clauses</i>			
Finite subordinate clauses	28.96 (8.48)	0.07 (0.01)	n.a.
Infinitive subordinate clauses	3.67 (2.67)	0.01 (0.01)	0.13 (0.09)*
Relative clauses	5.06 (2.98)	0.01 (0.01)	0.18 (0.09)*
<i>Noun phrase complexity</i>			
Length of noun phrases	3.06 (2.16)	n.a.	n.a.
<i>Passive voice</i>			
Number of passives	3.36 (2.48)	0.01 (0.01)	n.a.
<i>Words before main verb</i>			
Number of words before main verb	4.11 (0.58)	n.a.	n.a.

\*RF-ratio and IF-ratio

the Relative number of passives ( $t(123)=4.58$ ,  $p<0.001$ ), RRC ( $t(123)=2.02$ ,  $p<0.05$ ) and RF-ratio ( $t(123)=2.50$ ,  $p<0.05$ ) between texts arguing for or against a statement.

Mutual correlations between syntactic measures were explored by constructing a Pearson's correlation matrix. The only correlations stronger than 0.3 (which is considered the threshold for speaking of a weak correlation) were the correlations between Sentence length and Variation in sentence length (0.78), between the Relative number of infinitive clauses and the IF-ratio (0.91), and between RRC and RF-ratio (0.91). These correlations are not surprising given that the variables are partly based on the same measures. None of the syntactic measures had a significant direct correlation with text quality.

## Regression analysis

Table 4 shows the multilevel regression model with the syntactic measures as predictors and text quality as the outcome variables. Statement and Position were also included as predictors considering the fact that the previous analyses indicated an effect of these variables on some of the syntactic measures. Since there were high correlations between the Relative number of infinitive clauses and the IF-ratio (0.91) and between RRC and RF-ratio (0.91), and multicollinearity diagnostics of a preliminary regression model indicated multicollinearity issues for these variables, we performed a residual regression. In order to do so, we created variables that only contained the unique variance of the IF-ratio and the RF-ratio when the shared variance with the Relative number of infinite clauses and the RRC respectively was taken out. These variables were then used as predictors in the model instead of the original variables. As can be derived from the model, the significant predictors of text quality were the Relative number of finite subordinate clauses and the RF-ratio.

**Table 4** Regression model predicting text quality

	<i>b</i>	$\beta$	95% CI	Std. 95% CI	<i>p</i>
Intercept	-2.32	0.06	-6.15, 1.51	-0.40, 0.52	0.24
Statement	0.06	0.04	-0.56, 0.67	-0.35, 0.42	0.86
Position	-0.17	-0.11	-0.80, 0.45	-0.51, 0.29	0.59
Sentence length	-0.03	-0.13	-0.10, 0.04	-0.39, 0.13	0.34
Variation in sentence length	0.01	0.02	-0.10, 0.11	-0.25, 0.29	0.89
Finite subordinate clauses	56.59	0.53	17.12, 96.06	0.16, 0.90	0.005*
Infinitive subordinate clauses	31.23	0.12	-17.10, 79.57	-0.06, 0.29	0.63
Relative clauses (RRC)	8.29	0.04	-35.94, 52.53	-0.15, 0.22	0.71
IF-ratio (residual)	5.31	0.12	-6.00, 16.62	-0.14, 0.38	0.36
RF-ratio (residual)	21.71	0.53	9.69, 33.74	0.24, 0.83	0.001*
Length of noun phrases	-0.80	-0.13	-1.99, 0.40	-0.32, 0.06	0.19
Relative number of passives	-33.82	-0.13	-81.26, 13.63	-0.31, 0.05	0.16
Number of words before the main verb	0.21	0.08	-0.25, 0.67	-0.09, 0.25	0.37

\*Indicates significance at the  $p < 0.05$  level

Both significant predictors positively contributed to text quality, indicating that the quality of the text was generally higher when students used more finite subordinate clauses and when students had a higher RF-ratio (i.e., a higher number of relative clauses relative to the number of finite clauses). The marginal  $R^2$  value was 0.145, indicating that the predictors in our model explain 14.5% of the variance in text quality. The VIF values indicated no substantial problems with multicollinearity since they were all well below 10.

## Discussion

The current study explored the relation between measures of syntactic complexity and text quality in Dutch secondary students' L1 argumentative essays. To this end, 125 essays were analyzed on syntactic measures and rated on quality by means of comparative judgement (Pollit, 2012). A multilevel regression analysis was then carried out in order to investigate which measures of syntactic complexity predicted text quality. The results indicate two significant predictors, both positively contributing to text quality: the relative number of finite clauses and the ratio between the number of relative clauses and the number of finite clauses (RF-ratio).

The first finding, regarding the use of finite clauses, seems to contradict the previous literature. According to Myhill (2008), stronger writers prefer infinitive clauses over finite clauses. In our dataset, the opposite seems to hold: the more finite clauses a student uses, the better the quality of the resulting text. Using more infinitive clauses, on the other hand, does not result in better essay quality. This result might be explained by the fact that Myhill's (2008) texts were of a different genre than ours (personal narrative vs. argumentative essay on a social topic). According to Anbeek and Verhagen (2001) and Stukker and Verhagen (2019), using infinitive

clauses creates a sense of timelessness, which is more befitting of a narrative-like text than of an argumentative essay. In argumentative essays, expressing verb tense may be of importance in correctly interpreting the arguments presented. Hence, the different genres could account for the difference between our findings and Myhill's. In addition, certain sentence types may be more valued in some languages than in others, i.e., English readers may attribute more value to infinitive clauses in texts because the English language may simply attribute more value to them in general. According to Biber et al. (2002, p. 328), infinitive clauses in English are more common in written registers than in spoken registers, which may be an indication that infinitive clauses are considered typical for English writing, therefore contributing to their added value in English texts. Authoritative Dutch scientific grammars (e.g., Haeseryn et al., 1997; Broekhuis & Corver, 2015) or language advice books (e.g., Renkema, 2012) do not indicate such differences in usage between finite and non-finite clauses for Dutch, which may give clues about potential differences in usage between Dutch and English infinitive clauses. We leave these issues open for future research to explore.

The second positive predictor of text quality was the RF-ratio. Our results indicate that the higher the ratio between finite clauses and relative clauses is, the higher the quality of the resulting text is rated. This measure is related to variation in sentence patterning, a principle that was also found to be of importance by Myhill (2008). While Myhill (2008) shows that able writers vary sentence patterning by means of inversion, our results show that variation in sentence patterning might also be of consequence with regard to certain types of clauses. Interestingly, simply using more relative clauses per se does not lead to improved writing outcomes, but using them as a means to vary in sentence patterning does seem to have positive effects. Somehow, this effect stays limited to the RF-ratio, and it does not relate to the infinitive clauses—finite clauses ratio (IF-ratio), which yields a non-significant result. This suggests that not all variation in sentence patterning has equal effects on quality, which is further substantiated by the fact that variation in sentence length yields a non-significant result in the current study.

Why the RF-ratio has more predictive power than the IF-ratio might be explained by the fact that relative clauses seem to play a more important role in text cohesion (Boivin & Pinsonneault, 2008). In argumentative texts, relative clauses can be of particular importance since they are commonly used as expanding clauses, meaning that they elaborate on the information from their antecedent, or as restricting clauses, in which case they narrow the information expressed by the antecedent (Haeseryn et al., 1997; Halliday & Matthiessen, 2014, p. 493). It could well be that either expanding on or restricting the antecedent has a great effect on the quality of the argument. Thus, relative clauses can be used to convey more nuanced or fine-grained information, which benefits the overall coherence of the text (cf. Tol-Verkuy, 2005, pp. 190–193), or its argumentative power.

The fact that all of the other variables included in our study did not significantly predict text quality is somewhat surprising since previous work has found significant relations between these measures and text quality (e.g., Chall & Dale, 1995; Kameen, 1979; McNamara et al., 2010). Not only do these findings contradict earlier research findings, they also fail to provide support for the frequent

recommendation made by writing handbooks to avoid using passive constructions. Based on our results, no negative effect (or positive effect for that matter) on text quality can be expected from using passive constructions in essays. This does not mean that writers should suddenly cram their texts with passive constructions (for the sake of readability), but the advice to avoid them seems too strict in relation to the quality of the writing.

There are some possible explanations for the contradicting findings that have surfaced in our research. First, since our study was conducted in a Dutch context whereas most of the other studies cited in this paper are not, differences between these languages could play a role. Since different languages have different ways of expressing grammatical relations in texts (Verhoeven et al., 2002), distinct effects may arise due to these differences. The fact that our results are so different from previous work, suggests that there may be large differences between different languages, justifying more cross-linguistic research in examining relationships between syntactic features and text quality.

Secondly, grammatical features that were investigated in the present study may have been operationalized differently than in other studies (possibly due to language differences). For example, McNamara et al. (2010) find that the number of words before the main verb significantly predicts text quality, whereas our study did not find such an effect. Although McNamara et al. (2010) give an example of how a greater number of words before the main verb could affect text quality (p. 69), it remains somewhat unclear what is considered the main verb in their study because they do not give any examples with more than one verb. For example, they rightly state that sentences such as ‘she laughs’ are less complex (1 word before the main verb) than sentences such as ‘Thus, in syntactically simple English sentences there are few words before the main verb’ (7 words before the main verb). However, it is unclear whether they would consider the finite verb as the main verb, or the verb that conveys most information. In the sentence ‘In the garden, Jan must have elaborated on syntactic complexity’ the finite verb is *must*, and there are 4 words before the main verb; however, *elaborated* is the only autonomous verb, meaning that it alone can be the sentences predicate. If *elaborated* is chosen as the main verb (which in our view it should be), then there are 6 words before the main verb. Differences in operationalizing such measures could well have influenced the outcomes.

Thirdly, our study adopted an innovative way of measuring text quality (using comparative judgement), which is mostly seen as more reliable than other means of assessing the quality of texts (cf. Lesterhuis et al., 2016; Pollitt, 2012). Differences in the methods of establishing text quality could therefore have contributed to contradictory findings.

Fourthly, differences may also be attributed to genre (cf. Crossley, 2020; Verhoeven et al., 2002). For example, Myhill (2008) finds that good writers vary in sentence patterning, using short sentences alongside long ones to craft a specific rhetorical effect. In our study, variation in sentence length did not significantly predict text quality. However, as mentioned previously, the students’ texts investigated in Myhill (2008) were of a different nature than ours (personal narrative vs. argumentative essays on social topics). It may well be the case that variation in

sentence length as interpreted by Myhill (2008) plays a much larger role in narrative texts than in expository or argumentative essays.

Finally, another possible explanation for the contradictory findings could lie in the fact that the relations between measures of syntactic complexity and text quality might not be linear. It is very well imaginable that the relation between, for instance, the use of passive constructions and text quality does not simply take the form of the more (or the less) the better (i.e., a linear relation) but rather is one in which there is a certain optimum in the middle (i.e., a curvilinear relation). Such curvilinear relations have previously been found between aspects of the writing process and text quality (Vandermeulen, Van den Broek, Van Steendam & Rijlaarsdam, 2019). Future research on the link between syntactic complexity and text quality might also consider investigating non-linear relations in addition to linear relations in order to come to a more complete understanding of this relationship.

Future studies set in the Dutch context might also take other linguistic indices into account, such as lexical features or cohesive elements, especially in combination with the syntactic measures used in the current study, as interactions between various linguistic elements are likely to play a key role in influencing text quality (Crossley, 2020).

## Conclusion

The results of our study demonstrate that, in Dutch secondary school student's argumentative writing, the relative number of finite clauses and the ratio between the number of relative clauses and the number of finite clauses positively predict text quality. Furthermore, discrepancies between our results and those of previous studies indicate that the relations between syntactic feature and text quality may vary based on factors such as language and genre. More research, specifically focusing on the potential moderating effect of these factors, is needed to gain a clear picture on the syntactic features that play a crucial role in the quality of texts of varying nature. As the knowledge base grows, the acquired insights will be increasingly relevant for writing education, helping busy teachers focus their attention on those grammatical features that most strongly contribute to text quality.

**Acknowledgements** We wish to express our gratitude to Fontys University of Applied Sciences (Sittard) for providing some funding for conducting this research (without providing a grant). In addition, we wish to thank the following master's students of Fontys University of Applied Sciences (Sittard) for their help in analyzing the syntactic features: Janine Finster, Ruud Hermans, Hèlen Kiggen, Karlijn Maar, Jasmijn Mulkens, Ryanda Soomers en Judith Wijnhoven. Finally, we thank the anonymous reviewers of this paper, whose thoughtful comments have greatly improved this article.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Anbeek, T. & Verhagen, A. (2001). Over stijl [On style]. *Neerlandistiek.nl* 1(1) <http://www.neerlandistiek.nl/01.01/>.
- Benson, B., & Campbell, H. (2009). Assessment of student writing with curriculum-based measurement. In G. Troja (Ed.), *Instruction and assessment for struggling writers: Evidence-based practices* (pp. 337–353). London: The Guilford Press.
- Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Hillsdale: Lawrence Erlbaum Associates.
- Berman, R., & Verhoeven, L. (2002). Cross-linguistic perspectives on the development of text-production abilities: Speech and writing. *Written Language & Literacy*, 5(1), 1–43.
- Biber, D., Conrad, S., & Leech, G. (2002). *Longman student grammar of spoken and written english*. Harlow: Longman.
- 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.
- Boivin, M.-C., & Pinsonneault, R. (2008). *La grammaire moderne : description et éléments pour sa didactique*. Montréal, Canada: Beauchemin.
- Broekhuis, H., & Corver, N. (2015). *Syntax of Dutch: Verbs and verb phrases* (Vol. 2). Amsterdam: Amsterdam University Press.
- Broekhuis, H., & Keizer, E. (2012). *Syntax of Dutch, nouns and noun phrases*. Amsterdam: Amsterdam University Press.
- Chall, J., & Dale, E. (1995). *Readability revisited: The new Dale-Chall readability formula*. Cambridge: Brookline Books.
- Crossley, S. A. (2020). Linguistic features in writing quality and development: An overview. *Journal of Writing Research*, 11(3), 415–443.
- Crossley, S. A., Greenfield, J., & McNamara, D. (2008). Assessing text readability using cognitively based indices. *TESOL Quarterly*, 42(3), 475–493.
- Crossley, S. A., Weston, J., McLain Sullivan, S., & McNamara, D. (2011). The development of writing proficiency as a function of grade level: A linguistic analysis. *Written Communication*, 28(3), 282–311.
- Crowhurst, M. (1980) *The effect of syntactic complexity on writing quality: A review of research*. ERIC document number ED202024.
- Ekens, T., & Meestringa, T. (2013). *Beoordeling van en feedback op schrijven in de tweede fase [Evaluation of and feedback on writing in upepr secondary education]*. Enschede: SLO.
- Ferreira, F. (2003). The misinterpretation of non-canonical sentences. *Cognitive Psychology*, 47, 164–203.
- Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68(1), 1–76.
- Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology*, 99(3), 445–476.
- Haeseryn, W., et al. (Eds.). (1997). *Algemene Nederlandse Spraakkunst*. Groningen: Martinus Nijhoff.
- Halliday, M., & Matthiessen, C. (2014). *Halliday's introduction to functional grammar* (4th ed.). London: Arnold.
- Haswell, J. E. (1986). *Change in undergraduate and post-graduate writing performance: Quantified findings*. Technical report.



- Haswell, J. E. (1990). Applying the Perry and Belenky developmental schemes to English 101 students. Paper presented at the Wyoming Conference on English, Laramie.
- Hayes, J., & Flower, L. (1980). Identifying the organization of writing processes. In L. W. Gregg & E. R. Steinberg (Eds.), *Cognitive processes in writing* (pp. 3–30). Hillsdale: Lawrence Erlbaum.
- Henkens, L. (2010). *Het onderwijs in schrijven van teksten [Education in writing texts]*. Utrecht: Educational inspectorate.
- Hoogeveen, M. (2017). *Schrijfvaardigheid in het basisonderwijs: domeinbeschrijving ten behoeve van peilingsonderzoek [Writing in primary education: Domain descriptions for survey research]*. Enschede: SLO.
- Hudson, R. (2009). Measuring maturity. In R. Beard, D. Myhill, M. Nystrand, & J. Riley (Eds.), *SAGE handbook of writing development* (pp. 349–362). London: Sage.
- Kameen, P. T. (1979). Syntactic skill and ESL writing quality. In C. Yorio, K. Perkins, & J. Schachter (Eds.), *On TESOL '79: The learner in focus* (pp. 343–364). Washington, D.C.: TESOL.
- Klecan-Aker, J. S., & Hendrick, D. L. (1985). A study of the syntactic language skills of normal school-aged children. *Language, Speech, and Hearing Services in Schools*, 16(3), 187–198.
- Kleijn, S. (2018). *Clozing in on readability: How linguistic features affect and predict text comprehension and on-line processing (Doctoral dissertation)*. Utrecht: LOT.
- Kyle, K. (2016). *Measuring syntactic development in L2 writing: Fine grained indices of syntactic complexity and usage-based indices of syntactic sophistication (Doctoral dissertation)*.
- Lesterhuis, M., Van Daal, T., Van Gasse, R., Coertjens, L., Donche, V., & De Maeyer, S. (2018). When teachers compare argumentative texts: Decisions informed by multiple complex aspects of text quality. *Educational Studies in Language and Literature*, 18, 1–22.
- Lesterhuis, M., Verhavert, S., Coertjens, L., Donche, V., & De Maeyer, S. (2016). Comparative judgement as promising alternative to score competences. In G. Ion & E. Cano (Eds.), *Innovative Practices for higher education assessment and measurement* (pp. 120–140). Hershey: IGI Global.
- MacArthur, C., Graham, S., & Fitzgerald, J. (2016). *Handbook of writing research* (2nd ed.). New York: The Guilford Press.
- MacArthur, C. A., Jennings, A., & Philippakos, Z. A. (2019). Which linguistic features predict quality of argumentative writing for college basic writers, and how do those features change with instruction? *Reading and Writing*, 32(6), 1553–1574.
- Mak, W. M. (2001). *Processing relative clauses. Effects of pragmatic, semantic and syntactic variables. Doctoral dissertation*. Nijmegen: Radboud University.
- McNamara, D. S., Crossley, S. A., & Roscoe, R. D. (2013). Natural language processing in an intelligent writing strategy tutoring system. *Behavior Research Methods*, 45, 499–515.
- McNamara, D. S., Scottley, S. A., & McCarthy, P. M. (2010). Linguistic features of writing quality. *Written Communication*, 27(1), 57–86.
- Meijerink e.a. (2009). *Over de drempels met taal en rekenen: Hoofdrapport van de Expertgroep Doorlopende Leerlijnen Taal en Rekenen*. Enschede.
- Myhill, D. A. (2008). Towards a linguistic model of sentence development in writing. *Language and Education*, 22(5), 271–288.
- Myhill, D. A. (2009). From talking to writing: Linguistic development in writing. In: *Teaching and learning writing: Psychological aspects of education: Current trends: British Journal of Educational Psychology Monograph Series II* (6). British Psychological Society, Leicester, UK (pp. 27–44).
- Nezlek, J. B. (2008). An introduction to multilevel modeling for Social and Personality Psychology. *Social and Personality Psychology Compass*, 2, 842–860.
- Perin, D., & Lauterbach, M. (2016). Assessing text-based writing of low-skilled college students. *International Journal of Artificial Intelligence in Education*, 28(1), 56–78.
- Pitler, E., & Nenkova, A. (2008). Revisiting readability: A unified framework for predicting text quality. In: *Proceedings of the conference on empirical methods in natural language processing* (pp. 186–195). Association for Computational Linguistics.
- Pollitt, A. (2012). Comparative judgement for assessment. *International Journal of Technology and Design Education*, 22(2), 157–170.
- Renkema, J. (2004). *Introduction to discourse studies*. Amsterdam: John Benjamins.
- Renkema, J. (2012). *Schrijfwijzer*. Amsterdam: Boom.

- Sadler, D. R. (2009). Transforming holistic assessment and grading into a vehicle for complex learning. In G. Joughin (Ed.), *Assessment, learning and judgement in higher education* (pp. 1–19). Dordrecht: Springer.
- Sommet, N., & Morselli, D. (2017). Keep calm and learn multilevel logistic modeling: A simplified three-step procedure using Stata, R, Mplus, and SPSS. *International Review of Social Psychology*, 30(1), 203–218.
- Staphorsius, G. (1994). Leesbaarheid en leesvaardigheid. De ontwikkeling van een domeingericht meet-instrument. (Doctoral dissertation). Arnhem: Cito.
- Stukker, N. & Verhagen, A. (2019). *Stijl, taal en tekst. Stilistiek op taalkundige basis. [Style, language and text. Stilistics on a linguistic basis]*. Leiden: Leiden University Press.
- Til, A. van, Weerden, J. van, Hemker, B., & Keune, K. (2014). *Balans van de taalverzorging en grammatica in het basis- en speciaal basisonderwijs. Uitkomsten van de peiling in 2009 in groep 5, groep 8 en de eindgroep van het SBO*. Arnhem: Cito.
- Tol-Verkuyl, E. (2005). *Taal in tekst. Hoe je taal gebruikt om tekst vorm te geven [Language in text. How you use language to shape texts]*. Bussum: Coutinho.
- Van Silfhout, G. (2018). *Het schrijfonderwijs in primair en voortgezet onderwijs [Writing education in primary and secondary schools]*. Enschede: SLO.
- Vandermeulen, N., Van den Broek, B., Van Steendam, E., & Rijlaarsdam, G. (2019). In search of an effective source use pattern for writing argumentative and informative synthesis texts. *Reading and Writing*, 33, 239–266.
- Verhavent, S., De Maeyer, S., Donche, V., & Coertjens, L. (2017). Scale separation reliability: What does it mean in the context of comparative judgement. *Applied Psychological Measurement*. <https://doi.org/10.1177/0146621617748321>.
- Verhoeven, L., Aparici, M., Cahana-Amity, M., van Hell, J. V., Kriz, S., & Viguie-Simon, A. (2002). Clause packaging in writing and speech: A cross-linguistic developmental analysis. *Written Language and Literacy*, 5(2), 135–161.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Affiliations

Jimmy H. M. van Rijt<sup>1,2</sup>  · Brenda van den Broek<sup>3</sup> · Sven De Maeyer<sup>4</sup>

Brenda van den Broek  
brenda.vandenbroek@uantwerpen.be

Sven De Maeyer  
sven.demaeyer@uantwerpen.be

- <sup>1</sup> Department of Teacher Education, Fontys University of Applied Sciences, P.O. Box 558, 6130 AN Sittard, The Netherlands
- <sup>2</sup> Department of Linguistics, Center for Language Studies, Radboud University Nijmegen, P.O. Box 9103, 6500 HD Nijmegen, The Netherlands
- <sup>3</sup> Department of Linguistics, University of Antwerp, Prinsstraat 13, 2000 Antwerp, Belgium
- <sup>4</sup> Department of Educational Sciences, University of Antwerp, Sint-Jacobstraat 2-4, 2000 Antwerp, Belgium