

Refocusing writing style education? Relationships between stylistic lapses and the quality of Dutch secondary school students' argumentative texts

Jimmy H.M. van Rijt*, Brenda van den Broek[°] & Huub van den Bergh[#]

* Tilburg Center of the Learning Sciences, Tilburg University | the Netherlands

[°] Department of Linguistics, University of Antwerp | Belgium

[#] Utrecht Institute of Linguistics OTS, Utrecht University | the Netherlands

Abstract: In Dutch L1 classrooms, style in non-fictional genres is typically taught by means of normative exercises in which students are tasked to identify stylistic lapses. Not much is known about the effectiveness of such exercises when teaching style. Unknown factors include what kinds of stylistic shortcomings are found in Dutch students' writing, and how the occurrence of certain stylistic lapses relates to writing quality. The current study empirically explores these scarcely investigated issues. Teachers rated 125 argumentative texts written by tenth-grade pre-university students by means of comparative judgement. Additionally, these texts were manually analyzed to investigate the occurrence of stylistic lapses, taking into account stylistic lapses that are common in text books ('standard category') and other types of style related errors ('other category'). Multilevel regression analyses revealed that only one of the stylistic lapses from the standard category negatively influenced text quality as evaluated by teachers, namely the use of detached phrases. In the other category, only mistakes in question marks negatively predicted text quality. A final model including those two predictors explained 11.1% of the variance in text quality. The article discusses the implications of these findings for non-fictional style education, suggesting that it might need to be refocused.

Keywords: argumentative writing, style, stylistic lapses, text quality, comparative judgement



Van Rijt, J.H.M., Van den Broek, B., & Van den Bergh, H. (2021 - accepted for publication). Refocusing writing style education? Relationships between stylistic lapses and the quality of Dutch secondary school students' argumentative texts. *Journal of Writing Research*, x(x), ##-##. DOI: xx

Contact: Au Jimmy H.M. van Rijt, Tilburg Center of the Learning Sciences, Tilburg University, Postbus 90153, 5000 LE Tilburg | the Netherlands - j.h.m.vanrijt@tilburguniversity.edu

Copyright: Earli | This article is published under Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 Unported license.

1. Introduction

Learning to write entails learning how to put complex thoughts into words. With it comes the realization that there are many different ways to convey the same idea. Experienced writers know that depending on the intended audience of the text, the genre, and the writers' communicative goals, various writing styles may be adopted. Different styles can potentially convey different rhetorical effects, and could lead to slight differences in meaning. Some have even claimed that differences in style always lead to different meanings (e.g., Beardsley, 1967), although this claim is somewhat controversial (Stukker & Verhagen, 2019).

According to Renkema (2004, p. 145), style can be loosely defined as '[the possibility] to say the same thing in any number of different ways. The word *style* is used to denote these different ways.' However, Renkema also acknowledges that it is very difficult to adequately define style, in part because it can be viewed from different angles (e.g., 'style as a possible form for a specific content', 'style as a choice of specific patterns' and 'style as a deviation from expectations', Renkema 2004, p. 148-149). Stukker and Verhagen (2019) also acknowledge this difficulty, considering the dichotomy between dualistic ('style is a matter of form') and monistic ('style is a matter of meaning') views towards style as a central problem in defining the notion. An additional difficulty in understanding what *style* is, is that previous literature does not always clearly distinguish *style* from related notions such as *register* and *genre* (Lee, 2001). Biber and Conrad (2019, p. 15) consider *genre*, *register* and *style* as 'different approaches or perspectives for analyzing text varieties', meaning that every text can be studied from each of these perspectives. According to them, the *genre* perspective focuses on the linguistic characteristics that are used to structure complete texts, whereas *register* and *style* put more emphasis on 'the pervasive linguistic characteristics of representative text excerpts from the variety' (p. 15). *Register* then deals with functional variation within a text, whereas *style* is not about functional linguistic variation. Rather, Biber and Conrad (2019) maintain that using certain linguistic patterns is associated with aesthetic preferences, and that authors can thus have different views about the question what constitutes good style.

Research into style is a current topic in several disciplines. Style is, for example, often studied from the perspective of (classical) rhetoric, where it finds its roots, cf. Renkema (2004). In classical antiquity, *elocutio* was considered one of the key tasks of any orator. After *inventio* (discovery of arguments) and *dispositio* (organisation of arguments), *elocutio* was charged with the mastery of stylistic elements meant to enforce the argumentation that was employed. In this process, certain style virtues applied, such as using idiomatically 'pure' language (*puritas*) and choosing the appropriate and most effective forms of stylistic embellishment (*ornatus*). In essence, *elocutio* might be considered the teaching of style (Kennedy, 1994).

Currently, style is often studied from the perspective of *cognitive linguistics* (Stukker & Verhagen, 2019; Verhagen, 2012) and *stylistics* in general (Jeffries, & McIntyre, 2010; Macrae, 2016; Norgaard, Busse & Montoro, 2010). And even though style and stylistic choices are of interest in such branches of science once more, 'style scholarship has neglected to consider the role of classroom assessment in the teaching of style' (Medzerian, 2010, p. 187). While some stylisticians have paid attention to what has now become known as *pedagogical stylistics*, which is mainly concerned with 'the pedagogical usefulness and potential of stylistics for teaching (the language of) literature' (cf. Norgaard et al., 2010, p. 37), there is a lack of high quality research into non-literary or non-fictional style education.

Questions such as what style in non-fictional genres entails, how style in non-fictional genres is effectively taught and which learning activities are beneficial for this process are currently underresearched. Such questions cannot be answered adequately because of a lack of empirical (intervention) studies related to non-fictional style education (cf. Naciscione, 2010). Another important question that needs answering, is which aspects of style should be considered in educational settings, i.e., which content should be addressed in education, and how does this relate to writing quality? Our paper will address this matter from an L1 perspective.

1.1 Style in non-fictional writing

A lack of empirical studies into style education can be considered a crucial deficit in writing education research, since style is a key component in writing which typically 'appreciates the triadic relationship among reader, writer and text' (Medzerian, 2010, p. 201). In spite of an apparent lack of research into non-fictional style education, teachers appear to value writing style, since it is highly prioritized in teachers' commentary on written texts (Smith, 1997). In many cases, teachers' commentaries related to style are 'critical, with an underlying expectation of revision' (Otnes & Solheim, 2019). Given the significance that teachers attribute to style, it is important to determine the relevant content of non-fictional style lessons. Because writing style is co-determined by factors such as genre, goal and context, pertaining to both reader and writer, style should be viewed as a 'higher order feature' of the text. Lesterhuis et al. (2018) developed a coding scheme based on Cumming et al. (2001, 2002) to evaluate the quality of argumentative texts in which they considered style a complex, higher-order feature. Other features related to style— concepts such as *word use*, *language use*, *fluency* and *tone*— were also considered higher order aspects of texts. By contrast, 'usage' items, which Lesterhuis et al. (2018) dubbed 'mechanics', such as *spelling*, *capitalization*, *punctuation* and *grammatical errors*, were considered lower-order features of texts. This division in higher and lower features emphasizes the importance of such higher order aspects of texts, befitting a 'creativity discourse' of writing (Bomer, Land, Cira Rubin & Van Dike, 2019).

However, style is not viewed as a higher order aspect of the text by all accounts, and is certainly not treated as such in all educational settings. Medzerian (2010) states that some hold style-as-error views, which is ‘reflected in local, sentence-level treatments of style’ (p. 195). Renkema (2004, p. 150) calls this a *normative* approach to style in which style is predominantly taught in isolation. In this case, students are either presented with a generic set of style recommendations (e.g., ‘write short sentences’, ‘place yourself in the background’) or are presented with sentences containing style violations or *stylistic lapses*, which can be defined as ‘unintended deviations from a style type or style level required in a certain context’ (Claes & Hulsens, 2015). These include pleonasms (stylistic lapse in which part of the meaning of the word is repeated without a clear function), tautologies (stylistic lapse in which the full meaning of a word is expressed again in a different word which usually belongs to the same word class, without serving a reinforcing function), and contaminations (a mixture of two words or expressions). In some educational jurisdictions, such as the Netherlands, even grammatical errors are treated as if they were a matter of style. For example, many textbooks in secondary schools treat grammatical incongruences or agreement errors as a stylistic error, as we will elaborate on below. Students are then tasked to identify the kind of stylistic lapse they are confronted with, and they are stimulated to ‘correct’ the sentences.

In the Dutch secondary school context, in which the current study is conducted, identifying stylistic lapses appears to be the dominant approach to teaching style in non-fictional writing (Gerards & Van Rijt, 2018; Hulshof, 2009; Steenbakkens, 2018; Steenbakkens, Stukker & De Glopper, 2021). Such approaches to style education have been criticized for several reasons.

First, these isolated approaches strengthen the idea that style is detached from the overall writing process, and convey the inaccurate message that style should only be momentarily considered (Medzerian, 2010). This way, style cannot possibly be seen as a higher order aspect of the text, but instead turns into a mechanic. An additional problem is that research into writing processes also seems to focus little on the role of stylistic choices in students’ (argumentative) writing (cf. Van den Bergh, Rijlaarsdam & Van Steendam, 2016).

Second, stylistic lapse exercises nourish the false belief that style is an all-or-nothing phenomenon – style can either be completely good or bad; tropes and stylistic lapses can either be correct or false. This is, however, not the case. For example, pleonasms can be harmful to style in some contexts, but not in others (Lehmann, 2005). In a scenario where students are tasked to write a weather forecast, they could write something along the lines of: *Tomorrow, white snow will come down from the sky*. In this scenario, there is a limited need for the adjective ‘white’. This pleonasm could be considered harmful to the students’ style, because of the redundancy (snow is generally white, and the adjective serves no clear purpose here). On the other hand, it is not hard to imagine scenarios in which it is

beneficial to foreground the fact that the snow is white, for example in a context in which some snow is still white, whereas in other parts, the snow has turned to grey mush. Moreover, in writing a novel or a poem, the use of *white snow* might be considered stylistically sophisticated. Hence, genre also plays a role in style choices. Depending on context, genre and aim, experienced writers will either use a pleonasm or refrain from using it depending on whether they feel a literary function or concise writing is called for (Kashefi, Lukas & Hwa, 2018). Inexperienced writers do not yet possess the ability to make such conscious choices in their writing.

Since style is not an all-or-nothing phenomenon, but context-dependent and gradient, students should be taught how to reflect more consciously on the role of style and stylistic choices in their writing (Steenbakkens et al., 2021). Focusing mostly on isolated stylistic lapses does not facilitate such a reflective attitude towards style. Rather, as in the case of traditional parsing exercises, this tends to make learning writers more prereflective (King & Kitchener, 2004) in the sense that students are more focused on 'one correct answer' (cf. Wijnands, Van Rijt & Coppen, 2021) rather than on 'conscious writing proficiency' (Jansen, 2016; Gerards & Van Rijt, 2018), in which they can confidently weigh different style options.

1.2 Stylistic lapses in students' writing

In the Dutch context, secondary school text books typically cover style in the ninth and tenth grade, particularly in higher vocational education and pre-university education (Steenbakkens, 2018). In such lessons, eleven basic stylistic lapses are normally covered (see for example Frank et al., 2013): redundant forms of writing, including *pleonasms*, *tautologies* and *contaminations*; reference errors, including *vague or ambiguous* referencing, *norm violations in referencing* (e.g., using the pronoun *hen* instead of *hun* to refer to a group of people or vice versa); *agreement errors*; *incorrect contractions*; *incorrect inversion*; *erroneous non-finite clauses*, *detached phrases* and *dat-als constructions (that ... if constructions)*. (See Table 1 for an overview of these stylistic lapses, including examples from our data in Dutch with translations). As mentioned previously, several of these 'stylistic lapses' could be considered grammatical errors rather than stylistic deficiencies, pointing to normative tendencies in the teaching of style.

Steenbakkens (2018) claims that not all of these stylistic lapses occur in students' writing, which raises questions about whether or not it is important to pay attention to them in writing education. While agreement errors (Van de Gein, 2012), vague referencing (Van de Gein, 2012; Van Rijt, 2014) and detached phrases occur in texts regularly (Steenbakkens, 2018), stylistic lapses such as erroneous non-finite clauses, tautologies and pleonasms appear to occur very rarely in students' writing (Steenbakkens, 2018). In subsequent work, Steenbakkens et al. (2021) provide empirical evidence for this claim, showing that there are large frequency differences between various types of stylistic lapses. For instance, Steenbakkens et

al. (2021) find that pleonasms, tautologies and erroneous non-finite clauses are extremely rare in students' writing, whereas agreement errors occur much more often.

Other studies also show that students often tend to mix spoken and written language forms, or formal and informal language (errors related to register), they tend to write very long and confusing sentences (see also Myhill, 2008; Van de Gein, 2012) and they often punctuate in the wrong places (Steenbakkens, 2018; Van de Gein, 2012). However, there is a severe lack of research into the question what types of style related 'errors' students produce in their actual writing, nor has it been researched how such stylistic lapses relate to text quality.

While teachers spend a large amount of time on teaching style via isolated exercises, it is currently unknown whether such style exercises make sense. Do students actually engage in stylistic lapses when writing, and if so, how is this evaluated by their teachers? Does engaging in these stylistic lapses negatively affect how texts are rated? In particular, we aimed to answer the following research questions:

1. How often do the 'standard' stylistic lapses, which are covered in most standard Dutch secondary school text books, occur in students' argumentative writing?
2. Which other style-related violations do students engage in, and how often do these occur?
3. What is the relationship between both types of lapses ('standard' vs. 'other') on text quality as evaluated by Dutch language teachers?
4. What are the main consequences for style education?

The current study explores these questions in the Dutch secondary school context (10th grade, pre-university education), hereby deliberately focusing on the genre of argumentative writing, which is very common in Dutch secondary education. One specific genre is studied because of the interrelations between style and genre.

2. Method

2.1 Participants

Participants in this study were 136 tenth grade pre-university students (67 male, mean age 15.5) who were registered at five different Dutch secondary schools in the province of Limburg. One class of each school participated. Dutch (or a Dutch dialect) was the first language of 130 of those students, and ten of the participating students reported having a language disorder such as dyslexia. Classes were

selected for participation by their teachers, for which the schools gave their consent. No students were excluded from the dataset at this stage.

2.2 Materials

This study used an argumentative essay writing task¹ that was previously used successfully in another study (Van Rijt, Van den Broek & De Maeyer, 2021). Within this task, students could opt to argue in favour of or against one of the following 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 taken from the database of the Dutch debating institute (www.debatinstituut.nl) which includes validated statements suitable for debating. The statements were thus considered to be equally suited for an argumentative writing task. Both of the selected statements have a difficulty level of 2 out of 4, as determined 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, to avoid that students would copy and paste texts they found online. We deliberately did not allow the students the opportunity to revise their text, since it might be expected that more stylistic lapses can be found in unrevised texts than in revised texts.

2.3 Procedure

The participating students carried out the essay writing task on computers at their own secondary school. Before the actual writing task, participants completed a short online demographics questionnaire. Students received textual instruction on the writing task and had 50 minutes to finish their essays. They were encouraged to carry out the task as well as they could within this time frame and were requested to turn in a finished text.

2.4 Text quality rating

The text quality of 133 texts was rated holistically. One of the texts was not handed in correctly and proved untraceable, two of the texts were excluded because they were considered too short since they contained less than 200 words. The quality of the texts was rated by means of *D-PAC* (now called *Comproved* – see www.comproved.com): an online platform for comparative judgement. In comparative judgement, raters repeatedly compare two performances (in the case of the current study *texts*) and decide which of the two is better. Texts are compared multiple times to various other texts by multiple raters, resulting in a scale that ranks all the rated performances from worst to best (Lesterhuis, Verhavert, Coertjens, Donche and De Maeyer, 2016). Rating the quality of the texts this way has several principal advantages compared to other methods of text quality rating (Pollit, 2012) such as using rubrics as, for instance, employed by MacArthur et al. (2019) and

McNamara et al. (2010). It encourages raters to rate the quality of a text holistically instead of analytically, using a finite set of criteria relating to certain aspects of the text. Analytical rating has been criticized for being too constraining, potentially compromising the judgement's validity (Sadler, 2009). Lesterhuis et al. (2018) have shown that when argumentative texts are rated by means of comparative judgement, raters consider a broad range of aspects relevant to text quality. They conclude from this that comparative judgement is a valid way to assess the quality of (argumentative) texts, thereby corroborating the conclusion of Gill and Bramley (2013). Besides the advantage regarding validity, comparative judgement also eliminates complications resulting from sequential effects and differences in raters' focus (Pollitt, 2012).

The texts in this study were rated by 11 raters consisting of teacher trainers, teachers of Dutch enrolled in a Master of Education programme, and researchers in the field of linguistics or discourse studies. On average, they had seven years of experience in text assessment. Texts were compared 24.3 times on average. The amount of time a text is compared depends on how fast the software can reliably determine its place on the quality scale. The final reliability coefficient (SSR, or Scale-separation Reliability) was .86 which is considered high (cf. Verhavert, De Maeyer, Donche & Coertjens, 2017).

2.5 Analysis of stylistic lapses

We examined the occurrence of 'standard stylistic lapses' in the texts, as well as the occurrence of other stylistic norm violations (as recommended by Steenbakkers et al., 2021). To determine which types of 'other stylistic norm violations' were relevant, a random 10% of the corpus was investigated in-depth ($N=13$). In this process, all possible stylistic norm violations were examined and categorized bottom-up. If one error occurred in at least three texts, it was taken into account as a category. This bottom-up strategy was guided by some preliminary results from previous work (e.g., Van de Gein, 2012 – cf. Introduction section), and it deliberately adopted a broad perspective on issues that may touch upon style. This process thus resulted in several 'other norm violation-categories' (see Table 1 for details), namely elements related to wrong word choice (*inappropriate words, spoken language, and incorrect use of words*), punctuation (*wrong use of commas, mistakes in question marks, colon and semicolon mistakes*), article use (*expletive article use, erroneous article use*), preposition- and conjunction errors (p/c-errors) (*left-out p/c, erroneous p/c use, double prepositions*) and *derailed sentences* (sentence that is made up of two (or more) conflicting sentence patterns, resulting in an ungrammatical sentence). Errors in spelling, although occurring frequently, were considered beyond the scope of the current research. [Table 1: Appendix A.]

Based on the established list of stylistic lapses, the full corpus of texts was then stylistically analyzed by trained expert teams of educational masters' students ('coders') who already worked as in-service Dutch language teachers, with an average of 3.1 (*SD*2.3) years of teaching experience. Each group comprised of three or four coders. None of these teachers also acted as a rater of the texts. In total, four expert groups were formed randomly, two of which were trained in and focused on identifying the 'standard' stylistic lapses. The two other groups investigated the other stylistic norm violations.

To avoid inaccuracies in the analyses, all stylistic lapses (including those from the 'other' category) were given a working definition based on an authoritative Dutch language advice book (Renkema, 2012). Coders were instructed to only take note of a stylistic lapse if they felt as a Dutch language teacher that a particular case should be considered 'bad' style, or if the case could simply be said to be wrong (e.g., wrong punctuation). Thus, instances of style that could technically fall under a working definition, but that were not considered wrong, were not noted (think, for instance, about the previously discussed case on 'white snow', which would not be noted as a (wrong) pleonasm in all contexts).

Expert groups analyzed each text separately, and were only allowed to note a stylistic norm violation if at least 3 out of 4 or 3 out of 3 coders agreed that the alleged norm violation would fit into the working definitions that were used. If one of the coders felt that the other coders were wrong about an issue, the first author of the paper was consulted, who would cast a deciding vote. To maximize reliability, the expert groups also checked the texts that were analyzed by the other group. If they found any cases of doubt, the second group would discuss these points with the first group, and together, they would take a majority decision about how best to categorize a certain norm violation. Full agreement about the final coding decisions was reached after discussion.

In addition, the first author of the present paper checked a randomly selected 10% ($N=13$) of the texts in the corpus blindly for analysis accuracy at the end of the process, and found that overall, the expert groups had scored the texts according to the described procedure. Only one instance of a stylistic lapse was overlooked by the expert groups in this random sample, and no other deviations were found. Consequently, the analyses were deemed sufficiently reliable.

2.6 Statistical analysis

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, the data were explored. Next, correlations among stylistic lapses, correlations between the occurrence of stylistic lapses and text quality, and possible effects of statement (internet behavior or voting age) and position (in favour or against) on the stylistic lapses were examined. Finally, linear regression

analyses were carried out with text quality as the outcome measure and the occurrence of the stylistic lapses as predictors. Before performing these analyses the data were checked for collinearity and linearity of the relations between the outcome variable and the predictor variables. As neither collinearity nor nonlinearity appears to play a major role in our data, we summarized the results in Appendix B. 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 .10, indicating that between-school differences account for 10% of text quality differences. Nezlek (2008) considers even small ICC's as sufficient reason for multilevel modeling. Therefore, 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 stylistic lapses on text quality is the same for students from the various schools).

3. Results

3.1 Exploration of the data

Table 2 contains descriptive data showing how many students argued in favour of 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). There were no differences in the mean quality of the texts between statements ($t(123) = 0.08, p = .94$) or between positions ($t(123) = 0.69, p = .50$). Students wrote texts of 420 words on average ($SD = 79$). The number of words written did not differ based on the chosen statement ($t(123) = 1.17, p = .25$) or position they took (i.e., in favour of or against) ($t(123) = 0.03, p = .98$).

Table 2: Participants' distribution over positions and statements

Statement	N	In favour of	Against
A Internet behavior	72	11	61
B Voting age	53	28	25
Total	125	39	86

Table 3 summarizes the occurrences of the various stylistic lapses. For each stylistic lapse, an absolute N is listed, as well as the average occurrences of the stylistic lapse per text and per 100 words (as recommended by Biber and Conrad (2019, p. 62-64)). What stands out from Table 3, is that stylistic lapses or inadequacies from the 'other' category are much more frequent than those in the standard error category, with roughly a 2:1 ratio. In the standard error category, students tend to make the most

mistakes in referencing (N=343), either due to norm violations in referencing (N=145) or due to vague or ambiguous referencing (N=198). Students hardly engage in writing down erroneous non-finite clauses (N=3), nor do they use many pleonasms (N=9), tautologies (N=12), that ... if-constructions (N=15), incorrect inversions (N=20) or incorrect contractions (N=30).

In the 'other' category, students frequently experience problems related to register and word use. Incorrect use of words and using spoken language and inappropriate words, which can be conceptually linked, make up a total of 318 cases (24.7% of the 'other' category). Students also have difficulty punctuating, with wrong comma use being at the top of the list (N=295). If punctuation is extended to also include question mark mistakes and colon and semicolon mistakes, it amounts to 463 cases (35.95% of the 'other' category). Students also appear to have some difficulty in determining whether an article is required in a certain context or not (N= 273).

Appendix C provides a correlation matrix among the relative measures of the stylistic lapses. While several of the correlations are significant, none of the r -values exceed .35, indicating that the correlations are weak to moderate in strength. Several significant direct correlations were found between text quality and the various stylistic variables relative to text length. Significant negative correlations were found for two variables from the standard category (the relative amount of contaminations, $r = -.194$, $p = .031$, and the relative amount of detached phrases, $r = -.213$, $p = .017$), as well as for two variables from the 'other' category (the relative amount of wrongly used commas, $r = -.205$, $p = .022$, and the relative amount of question mark mistakes, $r = -.256$, $p = .004$).

Independent samples T-tests indicated that there were significant differences between statements and positions regarding the relative occurrence of some of the stylistic lapses. Between the chosen topics there were significant differences for norm violations in referencing ($t(113,109) = 2.245$, $p = .027$), as well as for incorrect use of words ($t(97,588) = -2.214$, $p = .029$) and abundant article use ($t(123) = -2,286$, $p < 0.01$). Between chosen positions there were also some significant differences, namely for norm violations in referencing ($t(121,027) = -2.499$, $p = .014$) and abundant use of articles ($t(123) = -.2024$, $p = .045$). Given these significant differences, statement and position were incorporated into the multilevel regression analyses.

3.2 Regression analyses

Three regression analyses were carried out in which we worked with the relative measures of the stylistic lapses since text length tends to be a strong predictor of text quality and analyses of other measures therefore need to take into account this relationship (Biber & Conrad, 2019; MacArthur, Jennings & Philippakos, 2019; Van Rijt, Van den Broek & De Maeyer, 2021).

Table 3: Occurrences of stylistic lapses (SD) sorted from uncommon to common

Stylistic lapse	Category (S/O)*	Absolute average per text (SD)	Average per 100 words (SD)	Absolute N (% of total)
Erroneous non-finite clauses	S	0.02 (0.20)	0.005 (0.04)	3 (0.15)
Pleonasm	S	0.07 (0.26)	0.02 (0.06)	9 (0.46)
Tautology	S	0.10 (0.30)	0.03 (0.08)	12 (0.61)
Dat/als-construction (that ... if)	S	0.12 (0.37)	0.03 (0.10)	15 (0.77)
Incorrect inversion	S	0.16 (0.41)	0.04 (0.10)	20 (1.03)
Incorrect contractions	S	0.24 (0.51)	0.06 (0.13)	30 (1.54)
Agreement errors	S	0.52 (0.78)	0.12 (0.18)	65 (3.33)
Contaminations	S	0.57 (0.87)	0.13 (0.21)	71 (3.64)
Detached phrases	S	0.75 (1.18)	0.19 (0.32)	94 (4.82)
Norm violations in referencing	S	1.16 (1.61)	0.28 (0.39)	145 (7.44)
Ambiguous/vague referencing	S	1.58 (1.96)	0.39 (0.52)	198 (10.15)
Total S		N/A	N/A	662 (33.95)
Double prepositions	O	0.09 (0.36)	0.02 (0.09)	11 (0.56)
Inappropriate words	O	0.09 (0.38)	0.02 (0.09)	11 (0.56)
Wrong article use	O	0.10 (0.32)	0.24 (0.08)	12 (0.62)
Left out p/c	O	0.22 (0.47)	0.05 (0.11)	27 (1.38)
Colon and semicolon mistakes	O	0.22 (0.63)	0.05 (0.15)	28 (1.44)
Derailed sentences	O	0.50 (0.80)	0.12 (0.19)	63 (3.23)
Erroneous p/c	O	0.97 (1.17)	0.23 (0.27)	121 (6.21)
Spoken language	O	0.98 (1.72)	0.22 (0.36)	123 (6.31)
Mistakes in question marks	O	1.12 (1.50)	0.28 (0.53)	140 (7.18)
Incorrect use of words	O	1.47 (1.77)	0.34 (0.41)	184 (9.44)
Abundant article use	O	2.18 (3.39)	0.51 (0.80)	273 (14.00)
Wrong use of commas	O	2.36 (2.60)	0.56 (0.60)	295 (15.13)
Total O		N/A	N/A	1288 (66.05)
Total of all stylistic inadequacies		N/A	N/A	1950 (100.00)

*Note S = standard errors; O = other errors

In all regression models, chosen statement (A/B) and position (in favour of/against) were also included as predictors since data exploration showed differences between statements and positions for several of the stylistic variables. Table 4 shows the first multilevel regression model, with the relative amounts of stylistic lapses from the standard category as predictors and text quality as the outcome variable. In the second model (Table 5), we only took the relative amount of stylistic lapses from the 'other' category as predictors into account. In the third model, finally, we combined the significant predictors from the previous two models in order to arrive at a final model (Table 6).

From Table 4, it can be inferred that the only standard stylistic lapse significantly predicting text quality is the relative number of detached phrases. This model, containing only the standard stylistic lapses as predictors, explains 11.7% of the variance in text quality. Table 5 shows that there is also one significant predictor from the other stylistic lapses category: mistakes in using question marks. This model, containing only the other stylistic lapses as predictors, explains 14.4% of the variance in text quality. Finally, Table 6 shows that the two significant predictors from the previous models remain significant when combined in one model. This final model explains 11.1% of the variance in text quality. A likelihood ratio test comparing the models in Table 4 and 5 to an than an intercept only model shows that these models do not provide a better fit to the data than the null model test ($\chi^2(16) = 6.98, p = .90$ and $\chi^2(17) = 18.87, p = .17$, respectively). This is of course a consequence of the relative large number of parameters that appear not to be related to text quality. The final model, as presented in Table 6, only takes into account the two significant predictors from the previous models. It seems to fit better to the data than the null model ($\chi^2(4) = 15.01, p < .01$), although such conclusions about a better fit should be interpreted with caution when using stepwise regression analysis (Henderson & Denison, 1989). Table 6 illustrates that distinct effects arise from using detached phrases and mistakes in questionmarks, with the latter being slightly more harmful to text quality than the former.

Table 4: Regression model predicting text quality for standard stylistic lapses (relative measures: number of lapses / number of words)

	estimate (95% CI)	standardized estimate (95% CI)	<i>p</i>
Intercept	0.62 (-0.19, 1.42)	0.10 (-0.33, 0.54)	.13
Statement	-0.08 (-0.67, 0.51)	-0.05 (-0.43, 0.32)	.79
Position	-0.18 (-0.80, 0.44)	-0.11 (-0.51, 0.28)	.57
Erroneous non-finite clauses	-129.83 (-758.30, 498.64)	-0.04 (-0.21, 0.14)	.69
Pleonasm	-11.94 (-463.71, 439.84)	-0.01 (-0.18, 0.17)	.96
Tautology	-268.70 (-596.69, 59.29)	-0.14 (-0.32, 0.03)	.11

Dat/als-construction (that ... if)	-30.66 (-308.50, 247.18)	-0.02 (-0.20, 0.16)	.83
Incorrect inversion	74.49 (-197.62, 346.60)	0.05 (-0.13, 0.22)	.59
Incorrect contractions	13.85 (-208.05, 235.22)	0.01 (-0.17, 0.19)	.90
Agreement errors	65.80 (-87.35, 218.95)	0.08 (-0.10, 0.25)	.40
Contaminations	-131.13 (-266.65, 4.38)	-0.17 (-0.35, 0.01)	.06
Detached phrases	-126.77 (-215.34, -38.20)	-0.25 (-0.43, -0.07)	.005*
Norm violations in referencing	-11.41 (-83.88, 61.06)	-0.03 (-0.21, -0.15)	.76
Ambiguous/vague referencing	16.29 (-36.29, 68.87)	0.05 (-0.12, 0.23)	.54

Note. * indicates significance at the $p < .05$ level.

Table 5: Regression model predicting text quality for other stylistic lapses (relative measures: number of lapses / number of words)

	estimate (95% CI)	standardized estimate (95% CI)	p
Intercept	1.18 (0.35, 2.01)	0.17 (-0.24, 0.57)	.005
Statement	-0.02 (-0.66, 0.61)	-0.02 (-0.42, 0.39)	.94
Position	-0.36 (-0.98, 0.26)	0.23 (-0.62, 0.16)	.25
Double prepositions	-8.64 (-307.30, 290.01)	-0.01 (-0.18, 0.17)	.96
Inappropriate words	-30.83 (-329.44, 267.79)	-0.02 (-0.19, 0.16)	.84
Wrong article use	-133.44 (-460.12, 193.23)	-0.07 (-0.24, 0.10)	.42
Left out p/c	-77.87 (-309.59, 153.85)	-0.06 (-0.23, 0.12)	.51
Colon and semicolon mistakes	40.87 (-152.59, 234.33)	0.04 (-0.15, 0.23)	.68
Derailed sentences	-46.75 (-184.70, 91.19)	-0.06 (-0.23, 0.11)	.51
Erroneous p/c	-32.03 (-132.27, 69.21)	-0.06 (-0.23, -0.12)	.54
Spoken language	-64.22 (-140.15, 11.72)	-0.15 (-0.32, 0.03)	.10
Mistakes in question marks	-101.52 (-183.20, -19.84)	-0.23 (-0.41, -0.04)	.015*
Incorrect use of words	-12.00 (-87.30, 63.29)	-0.03 (-0.23, -0.16)	.76
Abundant article use	8.02 (-28.47, 44.50)	0.04 (-0.15, 0.23)	.67
Wrong use of commas	-36.07 (-84.05, 11.90)	-0.14 (-0.32, 0.05)	.14

Note. * indicates significance at the $p < .05$ level.

Table 6: Regression model predicting text quality for significant predictors from previous models (relative measures: number of lapses / number of words)

	estimate (95% CI)	standardized estimate (95% CI)	<i>p</i>
Intercept	0.83 (0.07, 1.59)	0.17 (-0.28, 0.62)	.033
Statement	-0.12 (-0.69, 0.44)	-0.07 (-0.44, 0.28)	.67
Position	-0.33 (-0.93, 0.27)	-0.21 (-0.59, 0.17)	.28
Detached phrases	-95.71 (-180.71, -10.72)	-0.19 (-0.36, 0.02)	.027*
Mistakes in question marks	-101.14 (-176.70, -25.58)	-0.23 (-0.40, -0.06)	.009*

Note. * indicates significance at the $p < .05$ level.

4. Discussion

4.1 Summary of study objectives and findings

The current study set out to investigate which stylistic lapses occur in Dutch tenth grade pre-university students' argumentative writing and how they relate to the quality of the text (i.e., how they are evaluated by Dutch language teachers). To this end, the quality of 125 texts was evaluated using comparative judgement. Additionally, the stylistic lapses in the texts were manually analyzed by in-service teachers of Dutch language and literature. This analysis focused on eleven standard stylistic lapses that are typically taught in the Dutch curriculum ('standard category'), as well as on other stylistic shortcomings that could affect style in one way or another ('other category'). Results of multilevel regression analyses show that only two stylistic lapses could significantly predict text quality: the relative number of detached phrases (standard category) and the relative number of mistakes in question marks (other category).

4.2 Interpretation of main findings

While Dutch education on stylistic lapses tends to focus on eleven standard categories, students make a lot more errors in the 'other' category. This is an interesting finding, since this means that we might have to reconsider the amount of attention paid to these lapses in secondary education. Some of the standard stylistic lapses that teachers spend a fair amount of time on in class (e.g., pleonasms, tautologies, and incorrection contractions) rarely occur in students' writing. This finding is mirrored in data from another recent study into frequency of stylistic lapses (focusing on pleonasms, agreement, incorrect contractions, erroneous non-finite clauses and tautologies) in secondary school students' writing (Steenbakkens

et al., 2021). Moreover, these stylistic lapses do not appear to have a particularly negative effect on the overall text quality.

In investigating which stylistic shortcomings can be observed in students' writing, the current study has contributed empirical knowledge on language errors in Dutch secondary schools, which was seriously lacking (Van de Gein, 2012; Steenbakkens et al., 2021). Van de Gein (2012) conducted an exploratory analysis of language errors that occur regularly in Dutch secondary students' writing (although not focusing on stylistic lapses) and found that errors in referencing were the most salient category. This is mirrored in our data, in which reference related errors make up almost 18 percent of all the shortcomings that were found in the texts (the second largest category). One might consider it surprising that such style related errors have no bigger impact on the quality of the resulting text, since vague or ambiguous referencing can seriously affect a text's overall coherence (Karremans & Van Enschoot, 2013). However, it is likely that the mere fact that such errors occur so often, causes them to no longer be a reason for text raters to see such errors as a factor separating good texts from bad texts.

Detached phrases and mistakes in the use of question marks do appear to affect text quality. In the case of the former, raters may feel that detached phrases are a sign of an incomplete understanding of how sentences are formed and linked together, which can be seen as a very basic understanding of how writing works. After all, students should have mastered such basic writing at this stage (grade 9/10 of secondary pre-university education), as is articulated in formal documents (Werkgroep Taal/Expertgroep Doorlopende Leerlijnen Taal en Rekenen, 2008). Even though it is not always wrong to use detached phrases (they can be used to convey certain rhetorical effects), Dutch language teachers do not appear to appreciate students' use of such constructions, especially because in the cases found in our data, no rhetorical effects appear to have been pursued by the students. Alternatively, students' unsuccessful use of detached phrases might be indicative of an attempt to use such forms for rhetorical effect. In the case of making errors in question marks, a similar argument could be made. Typically, question marks are considered very easy aspects of punctuation, which are first covered at the very early stages of primary education (Werkgroep Taal/Expertgroep Doorlopende Leerlijnen Taal en Rekenen, 2008). Hence, teachers are likely to feel that errors of such nature should no longer occur in the later stages of secondary education, in particular at pre-university level.

4.3 Implications

Looking at the two categories that were distinguished in the current study, it can be observed that the 'other' category contains relatively simpler notions (e.g., punctuation, article use, etc.). The standard errors that are covered in language education tend to be more conceptually advanced (e.g., pleonasms, tautologies). It

might be argued that the very fact that standard stylistic lapses occur very rarely can be attributed to the effectiveness of current style education. This seems an unlikely explanation, however, given that conceptually simpler mistakes occur much more frequently in spite of the many hours of education on these topics. Alternatively, the setting in which this writing task was performed may have led students to produce relatively simple sentences, thereby avoiding many stylistic errors. Future studies should control for this.

The fact that students make so many mistakes in the 'other' category, implies that it might be wise to attribute more attention to less advanced stylistic or style-related indices, such as punctuating properly, as well. This is not to say that more advanced matters of style should be deemed irrelevant in language education (indeed, it can be quite useful to know about pleonasms or tautologies), but it appears that educational practice would benefit from changes in non-fictional style education in two important ways.

First, more 'groundwork' appears to be needed before students are ready to understand more advanced stylistic issues. It would make sense to make sure students avoid the more basic errors (from the 'other category') first before they are ready to move on to the more sophisticated stylistic lapses (from the 'standard category') in education, especially given the 2:1 ratio found in the current study. Attention to the basic errors should be given throughout the curriculum, even in stages when students are already expected to fully master them.

Second, the way in which education in non-fictional style is shaped in the Dutch context is currently much more geared towards seeing style as a mechanic or a lower order feature of the text rather than towards seeing style as a higher order feature of the text (Lesterhuis et al., 2018).² This is expressed in the normative exercises that students are presented with in style education. The results of the current study suggest that text raters (i.e., Dutch language teachers) are hardly bothered by the occurrence of standard stylistic lapses (if these are present in the text at all). This raises the question whether it would not be more fruitful to approach style in a less 'mechanical' way. In other words: rather than teaching students how to identify stylistic lapses in isolation, style education might spend more attention to style in relation to the whole text. This way, students experience that style is really not about avoiding errors (i.e., style as a mechanical, lower-order application), but rather, that style can have a profound influence on the text as a whole (i.e., style as a higher order feature). In order to facilitate such a shift in focus, style should be addressed within the context of writing (instead of being rather isolated), and students should be invited to reflect consciously on their stylistic choices (Jansen, 2016; Steenbakkens, 2018, 2021). This could be done by teaching students how style is treated from the perspective of (classical) rhetoric, for example – one exploratory study has shown that this can lead to much more conscious reflection on the use of style in writing argumentative essays (Gerards &

Van Rijt, 2018). Such a focus shift does not have to be at odds with our previous recommendation to also spend more time on improving more basic aspects of students' writing (e.g., question mark use or using detached phrases).

On a separate note, the linguistic underpinning of some of the standard stylistic lapses is conceptually questionable, as is also pointed out by Steenbakkens et al. (2021). Not all of the standard stylistic lapses are sufficiently linguistically motivated to be taught as a 'rule' at school. In fact, some of these lapses are presented in commonly used text books as being wrong, while this is in fact not (always) the case. This has been argued, for example, in the case of 'erroneous' non-finite clauses (Coppen, 2006; Haeseryn et al., 1997) as well as for variants of incorrect contractions (Renkema, 2012) and some forms of agreement errors (Haeseryn et al., 1997; Renkema, 2012), particularly in cases of agreement in special partitive constructions (Hogeweg, Ramachers & De Hoop, 2018). This means that policy makers, teacher educators and particularly text book writers would do well to re-evaluate the content of some of the standard stylistic lapses that are covered in language education, so they can make scientifically sound decisions on which stylistic content should be taught. Particularly, the discrepancies between stylistic norms in language advice books and those in school books should be resolved. Steenbakkens et al. (2021) show that style related norms in school text books are much more rigid than style related norms as presented by language advice books or websites. In many instances, school books will consider a certain style related element as 'wrong', whereas scientifically underpinned language advice books are much more forgiving. We agree with Steenbakkens et al. (2021) that education should prepare students for learning how to consciously reflect on stylistic choices rather than pretending that there is an absolute norm when it comes to style.

Education might draw on insights from phraseology for more scientifically informed style education (cf. Gragner & Meunier, 2008; Hulshof, 2009), or it can draw on linguistically underpinned checklists for assessing style in various text genres, such as Stukker and Verhagen's (2019) checklist for evaluating stylistic choices. Future research might explore the potential of such approaches and instruments.

5. Limitations and future research

Although the current study provides relevant insights regarding the relationship between stylistic lapses and the quality of argumentative writing (particularly in the Dutch context), the study is not without limitations. Firstly, it is notoriously difficult to determine objectively what constitutes a stylistic lapse. When can a stylistic lapse be considered 'wrong'? The answer to this question is not obvious, because many cases are context-dependent. Of course, some stylistic lapses or style related errors are obviously either wrong or correct, but for the majority of stylistic issues discussed, it is more difficult to arrive at a decision. In order to optimize objectivity,

the evaluating teachers involved had to reach a majority decision on whether they felt something constituted a specific stylistic lapse. However, given the nature of some of the stylistic lapses, it is not unthinkable that in some cases different teachers would have made different choices, which could have led to different outcomes. In addition, our method of determining stylistic lapses (based on majority decisions) might result in underidentification of stylistic lapses. Future studies may address this issue by having the texts coded by independent raters. Another disadvantage of working with Dutch language teachers is that not all of them may be equally skilled at identifying some of the more challenging stylistic lapses, such as incorrect inversions, in spite of their training. On the other hand, the advantage is that judgements based on teachers contribute to ecological validity. Moreover, other steps have been taken to ensure that the analyses were carried out as sound as possible (e.g., a random check of the sample by the principal investigator and the consensus requirement). Other studies would do well to pay more specific attention to interrater reliability, although this will be challenging due to the nature of evaluating stylistic choices.

In addition, it may be argued that the outcomes of this study have been affected by our application of comparative judgement, rather than basing text quality on a rubric – a more common way for Dutch language teachers to evaluate texts. In our mind, the methodological advantages of using comparative judgement (see Pollitt, 2012; Lesterhuis et al., 2016) outweigh the potential advantage of using a more common way of assessing text quality (e.g., by means of a rubric). In addition, while comparative judgement is different from teachers' regular ways of assessing, Lesterhuis et al. (2018) have shown that comparative judgement encourages teachers to make decisions on complex and multiple aspects of text quality, in much the same way as they would when using a rubric or other methods of assessing. While more research is needed on comparing comparative judgement with other methods of evaluation (Lesterhuis et al., 2018), it seems likely that our application of comparative judgement will not have had a major impact on the results.

It was beyond the scope of the current paper to present more qualitative analyses to illustrate the effect of certain stylistic choices on writing quality. Future research would benefit from also taking such a qualitative approach. Moreover, it would be very interesting to gain a deeper understanding of how students deal with style within their writing processes, for example by investigating their thinking in think aloud conditions. Hopefully, the current study will serve as a starting point for more research on style education in non-fictional genres.

Acknowledgements

We wish to express our gratitude to prof. dr. Hans Hulshof (Leiden University) and prof. dr. Debra Myhill (University of Exeter) for their pre-review comments on this

manuscript, as well as to the editor and anonymous reviewers, who helped to improve this paper significantly. Finally, we are grateful to the following Master's students of Fontys University of Applied Sciences (Sittard) for their help in stylistically analyzing the students' texts: Sonja Alers, Pim van Basten-Batenburg, Joost Bremen, Marijke Eggen, Leonie Goutier, Vio Heuts, Lynne Janssen, Celine Leblanc, Mieke van Leeuwen, Silvia Oostindie, Bente Pieters, France van Rijn, Lotte Vennekens and Daniëlle Verlaar.

Notes

1. Given the exploratory nature of this study, we did not intend to gain a 'complete' understanding of the participating students' writing competence. Gaining a more or less complete understanding of students' writing ability would require them to write more texts (see e.g., Breetvelt, Van den Bergh & Rijlaarsdam, 1994).
2. In some cases, the current study has therefore also adopted a lower-order stance towards style. Counting stylistic lapses (and style-related deficiencies) in a way facilitates the perception of style as a lower-order aspect of a text.

References

- Beardsley, Monroe C. (1967) Style and Good Style. In: Martin Steinmann Jr. (Ed.), *New Rhetorics*. New York: Scribner's, pp. 191-213.
- Biber, D. & Conrad, S. (2019). *Register, Genre, and Style*. Second edition. Cambridge: Cambridge University Press.
- Bomer, R., Land, C., Cira Rubin, J. & Van Dike, L. (2019). Constructs of teaching writing in research about literacy teacher education. *Journal of Literacy Research* 51(2), 196-213. <https://doi.org/10.1177/1086296X19833783>
- Breetvelt, I., Van den Bergh, H. & Rijlaarsdam, G. (1994). Relations between Writing Processes and Text Quality: When and How? *Cognition and Instruction* 12(2), 103-123. https://doi.org/10.1207/s1532690xci1202_2
- Claes, P. & Hulsens, E. (2015). *Groot retorisch woordenboek: lexicon van stijlfiguren*. [Great rhetorical dictionary: lexicon of tropes]. Nijmegen: VanTilt.
- Coppen, P.-A. (2006). Strikt genomen. [Striktly speaking.] Blogpost. Consulted from <http://taalprof.blogspot.com/2006/10/strikt-genomen.html> on February 2nd, 2020.
- Cumming, A., Kantor, R., & Powers, D. (2001). *Scoring TOEFL essays and TOEFL 2000 prototype writing tasks: An investigation into raters' decision making and development of a preliminary analytic frame-work*. Ewing, NJ: Educational Testing Service.
- Cumming, A., Kantor, R., & Powers, D. (2002). Decision making while rating ESL/EFL writing tasks: A descriptive framework. *The Modern Language Journal* 86(1), 67-96. <https://doi.org/10.1111/1540-4781.00137>
- Field, A. (2013). *Discovering statistics using IBM SPSS Statistics* (4th edition). Washington: Sage.
- Frank, H., E. Grezel, M. Kooiman, W. Steenbergen & H. Zetstra (2013). *Nieuw Nederlands 4-vwo*. [New Dutch 4 vwo]. Groningen: Noordhoff.
- Gill, T., & Bramley, T. (2013). How accurate are examiners' holistic judgements of script quality? *Assessment in Education: Principles, Policy & Practice* 20(3), 308-324. <https://doi.org/10.1080/0969594X.2013.779229>
- Gragner, S. & Meunier, F. (2008). *Phraseology. An interdisciplinary perspective*. Amsterdam: John Publishing. <https://doi.org/10.1075/z.139>

- Otnes, H. & Solheim, R. (2019) Acts of responding. Teachers' written comments and students' text revisions. *Assessment in Education: Principles, Policy & Practice* 26(6), 700-720. <https://doi.org/0969594X.2019.1595524>
- Gerards, L. & Van Rijt, J. (2018). Bewuste taalvaardigheid bewerkstelligen met behulp van de klassieke retorica. Een didactische verkenning. [*Towards conscious language proficiency with the help of classical rhetoric. A pedagogical exploration.*] *Internationale neerlandistiek* 56(1), 1-22. <https://doi.org/10.5117/IN2018.1.GERA>
- Haeseryn, W. et al. (Eds.).(1997). *Algemene Nederlandse Spraakkunst. [General Dutch Reference Grammar]*. Groningen: Martinus Nijhoff.
- Henderson, D.A. & Denison, D.R. (1989). Stepwise Regression in Social and Psychological Research. *Psychological Reports* 64, 251-257. <https://doi.org/10.2466/pr0.1989.64.1.251>
- Hogeweg, L., Ramachers, S. & De Hoop, H. (2018). Singular Agreement in Special Partitive Constructions in Dutch. *Journal of Germanic Linguistics* 30(4), 335-370. <https://doi.org/10.1017/S1470542717000149>
- Hulshof, H. (2009). Holistische en analytische principes in taalkunde en taalonderwijs. [Holistic and analytical principles in linguistics and language education.] In R. Boogaart et al. (Eds.) *Woorden wisselen. Voor Ariane van Santen bij haar afscheid van de Leidse universiteit*, pp. 285-299. Leiden: Stichting Neerlandistiek Leiden.
- Jansen, C. (2016). Taal-en communicatieonderwijs in beweging. [*Language and communication education on the move*]. *Internationale Neerlandistiek* 54(2), 137-146. <https://doi.org/10.5117/IN2016.2.JANS>
- Jeffries, L. & McIntyre, D. (2010). *Stylistics. Cambridge Textbooks in Linguistics*. Cambridge: Cambridge University Press.
- Karremans, J. & Enschot, R. van (Eds.) (2013). *Tekstanalyse. Methode en toepassingen. [Text analysis. Method and applications]*. Assen: Van Gorcum.
- Kashefi, O., Lucas, A. & Hwa, R. (2018). Semantic Pleonasm Detection. *Proceedings of NAACL-HLT*, 225-230. <https://doi.org/10.18653/v1/N18-2036>
- Kennedy, G. (1994). *A new history of classical rhetoric*. Princeton: Princeton University Press.
- King, P.M. & Kitchener, K.S. (2004). Reflective judgement: Theory and research on the development of epistemic assumptions through adulthood. *Educational Psychologist* 39(1), 5-18. https://doi.org/10.1207/s15326985ep3901_2
- Lee, D. (2001). Genres, registers, text types, domains and styles: clarifying the concepts and navigating a path through the BNC jungle. *Language Learning and Technology* 5(3), 37-72. https://doi.org/10.1163/9789004334236_021
- Lehmann, C. (2005). "Pleonasm and hypercharacterisation", in: Geert Booij – Marle van Jaap (Eds.), pp.119-154. Markus, Manfred. https://doi.org/10.1007/1-4020-4066-0_5
- Lesterhuis, M., Verhavert, S., Coertjens, L, Donche, V., & De Maeyer, S. (2016). Comparative Judgement as Promising Alternative to Score Competences. In G. Ion and E. Cano (Eds.), *Innovative Practices for higher education assessment and measurement* (pp. 120-140). Hershey, PA: IGI Global.
- 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. *L1-Educational Studies in Language and Literature* 18, 1-22. <https://doi.org/10.17239/L1ESLL-2018.18.01.02>
- 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. <https://doi.org/10.1007/s11145-018-9853-6>
- Macrae, A. (2016). Stylistics. In: M. Giovanelli & D. Clayton (Eds.) *Knowing about language: Linguistics in the secondary English classroom*. London: Routledge, pp. 51-63.
- Mason, C.H. & Perreault, W.D. (1991). Collinearity, Power, and Interpretation of Multiple Regression Analysis. *Journal of Marketing Research*, 28(3), 268-280.

- <https://doi.org/10.2307/3172863>
- Medzerian, S. (2010). Style and the Pedagogy of Response. *Rhetoric Review* 29(2), 186-202. <https://doi.org/10.1080/07350191003613468>
- Myhill, D. A. (2008). Towards a Linguistic Model of Sentence Development in Writing. *Language and Education* 22(5) 271-288. <https://doi.org/10.2167/le775.0>
- Naciscione, A. (2010). Stylistic Use of Phraseological Units in Discourse. Amsterdam: John Benjamins. <https://doi.org/10.1075/z.159>
- Nezlek, J. B. (2008). An introduction to multilevel modeling for Social and Personality Psychology. *Social and Personality Psychology Compass* 2, 842-860. <https://doi.org/10.1111/j.1751-9004.2007.00059.x>
- Norgaard, N., Busse, B. & Montoro, R. (2010). *Key terms in Stylistics*. London: Continuum International Publishing Group.
- Pollitt, A. (2012). Comparative judgement for assessment. *International Journal of Technology and Design Education* 22(2), 157-170. <https://doi.org/10.1007/s10798-011-9189-x>
- Renkema, J. (2004). *Introduction to Discourse Studies*. Amsterdam: John Benjamins. <https://doi.org/10.1075/z.124>
- Renkema, J. (2012). *Schrijfwijzer. [Writing guide]*. 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, The Netherlands: Springer. https://doi.org/10.1007/978-1-4020-8905-3_4
- Smith, S. (1997) The Genre of the End Comment: Conventions in Teacher Responses to Student Writing. *College Composition and Communication* 48(2): 249-68. <https://doi.org/10.2307/358669>
- 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. <http://doi.org/10.5334/irsp.90>
- Steenbakkens, J. (2018). Stijlperspectieven – nieuwe lessen schrijfstijl voor midden- en bovenbouw vo. [*Style perspectives – new lessons in writing style for middle and higher secondary school classes*]. In: S. Vanhoooren & A. Mottart (Eds.) *Tweeëndertigste Conferentie Onderwijs Nederlands*, pp. 281-285. Gent: Academia Press.
- Steenbakkens, J. (2020). Reflecteren op de effecten van schrijfstijl. [*Reflecting on the effects of style*]. In: WODN Werkgroep Onderzoek Didactiek Nederlands (Ed.), *Handboek Didactiek Nederlands. Levende Talen – Zo kan het ook*. Consulted April 12, 2021 via <https://didactieknederlands.nl/zokanhetook/2021/03/reflecteren-op-de-effecten-van-schrijfstijl/>.
- Steenbakkens, J., Stukker, N. & De Glopper, K. (2021). Formuleerproblemen in leerlingenwerk en schoolboeken Nederlands: een onderzoek naar de aansluiting tussen de leerstof en de formuleerproblemen van leerlingen. [*Style problems in students' writing and in text books for Dutch: a study into the connection between the learning materials and students' problems in formulation*]. *Pedagogische Studiën* 98(1), 27-45.
- Stukker, N. & Verhagen, A. (2019). *Stijl, taal en tekst: stilistiek op taalkundige basis. [Style, language and text: stylistics on linguistic basis]*. Leiden: Leiden University Press.
- Van de Gein, J. (2012). Taal op klompen. Uitkomsten van een onderzoek naar taalfouten aan het einde van het voortgezet onderwijs. [*Results of a study into language errors at the end of secondary education*]. *Levende Talen Tijdschrift* 13(4), 22-30.
- Van den Bergh, H., Rijlaarsdam, G., & van Steendam, E. (2016). Writing process theory: A functional dynamic approach. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (p. 57–71). The Guilford Press.
- Van Rijt, J. (2014). De hypothetische meelezer: over de effecten van een reader-oriented viewdidactiek op het verwijfswoordengebruik van leerlingen uit 4-vwo. [*The hypothetical co-reader: on the effects of a reader-oriented view approach on pre-university students' use of reference words*]. *Levende Talen Tijdschrift* 15(4), 15-25.

- Van Rijt, J., Van den Broek, B. & De Maeyer, S. (2021). Syntactic predictors for text quality in Dutch upper-secondary school students' L1 argumentative writing. *Reading and Writing* 34, 449-465. <https://doi.org/10.1007/s11145-020-10079-5>
- Verhagen, A. (2012). Construal and Stylistics – within a language, across contexts, across languages. In: *Stylistics across Disciplines*. Conference Proceedings. Leiden. 1-16.
- Verhavert, 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>.
- Werkgroep Taal / Expertgroep Doorlopende Leerlijnen Taal en Rekenen (2008). *Over de drempels met taal: De niveaus voor de taalvaardigheid [Over the thresholds with language: The levels for language proficiency]*. Enschede: SLO.
- Wijnands, A., Van Rijt, J. & Coppen, P.-A. (2021). Learning to think about language step by step: a pedagogical template for the development of cognitive and reflective thinking skills in L1 grammar education. *Language Awareness*. Advance online publication. <https://doi.org/10.1080/09658416.2021.1871911>

Appendix A

Table 1: Stylistic lapses with definitions and examples from the data

Stylistic lapse (standard) (Dutch translation)	Definition based on Renkema (2012)*	Example from the data (with translations**) underlinings indicate the stylistic lapse
Pleonasm (Pleonasme)	A stylistic lapse in which part of the meaning of the word is repeated without a clear function.	(...) want ik vind dat iedereen recht heeft <u>op zijn eigen privacy</u> . '(...) because I feel that everyone has a right to his own privacy'
Tautology (Tautologie)	A stylistic lapse in which the full meaning of a word is expressed again in a different word which usually belongs to the same word class, without serving a reinforcing function.	<u>Dus al met al</u> vind ik 'Overall in conclusion I feel ...'
Contamination (Contaminatie)	A mixture of two words or expressions.	(...) ook zal je kind het op een bepaalde leeftijd niet meer pikken als je altijd <u>over zijn vingers meekijkt</u> 'your child will also no longer accept it at a certain age if you always look over his fingers' (contaminated 'op de vingers kijken' and 'meekijken')
Vague referencing (Vage verwijzing)	A reference to a referent or antecedent that has not been previously introduced, but appears to be present in the mental representation of the writer, or general unclarity in terms of what is referred to (Van Rijt, 2014)	Misschien stemmen ze wel op de zelfde als hun ouders, gewoon omdat ze te lui zijn om <u>het</u> zelf te onderzoeken. 'Maybe they vote for the same as their parents, simply because they are too lazy to investigate <i>it</i>
Ambiguous referencing (Ambigue verwijzing)	Using a pronoun that can indicate two or more	Zitten zij te appen, spelletjes te spelen of doen ze heel iets anders en wordt dit online-

	possible referents. (Van Rijt, 2014).	gedrag ook gecontroleerd door bijvoorbeeld een ouderrol? Naar mijn mening wordt <u>dit</u> te weinig gedaan 'Are they texting, playing games or doing something completely different and is this online behavior also controlled by, for example, a parent role? In my opinion <i>this</i> is not done enough' (<i>dit (this)</i> might refer to 'gecontroleerd' ('controlled') or to 'appen ... anders ('texting ... different').
Normviolations in referencing (Normschending in verwijzing)	Using a different grammatical gender than the norm prescribes in referencing, or using otherwise wrong pronouns to refer to an antecedent (e.g., using <i>hun</i> as a personal pronoun instead of <i>hen</i> , or vice versa).	Ook wordt er naar mijn mening geen rekening gehouden met <u>de jeugd</u> en <u>hun</u> behoeftes. 'Also, in my opinion, youth and their needs are not taken into account' (<i>youth</i> is grammatically singular, and <i>hun</i> is grammatically plural).
Agreementerrors (Congruentiefout)	Using singular forms of the finite verb when plural forms are called for and vice versa.	Ten tweede denk ik dat <u>je</u> je kind moeilijker met dingen <u>kunnen</u> vertrouwen (...) 'Secondly, I think it is harder for you to trust your child with things' (je (you) = singular; kunnen (no 1-to-1 translation) = plural).
Incorrect contractions (Foutieve samentrekking)	A style figure in which a verb or adjective connects two other word groups, while that word fulfills a different function in relation to both phrases.	Het blijft natuurlijk wel een gevaarlijke omgeving voor sommige kinderen <u>die</u> heel makkelijk beïnvloed kunnen worden en [...] zo hele nare dingen overkomen. 'Of course, it remains a dangerous environment for some children who can be influenced

Incorrect inversion (Foutieve inversie)	Style figure when coordinating main sentences, where the subject of the second main clause incorrectly follows the finite verb.	very easily and thus come across very nasty things.’ (<i>die</i> is the subject in the first sentence, but in the contraction it serves as the indirect object). Doordat ze nog niet de consequenties kunnen inzien en dat <i>moeten ze</i> goed leren (...) ‘Because they cannot yet see the consequences and they have to learn that well’
Erroneous non-finite clauses (Foutief beknopte bijzin)	The left-out subject of an infinitive clause must be equal to the subject of the main clause.	<i>Kijkend naar de jongens</i> zoeken ze vaak dingen op als porno ‘Watching the boys, they often look up things like porn’ (Sentence can be read to mean that the boys are watching themselves rather than someone else, which is intended).
Detached phrases (Losstaand zinsgedeelte)	A subordinate clause that is detached from the main sentence.	Op deze manier kunnen kinderen meestemmen op de zaken die effect op ze zullen hebben. <i>Terwijl de ouders de mogelijkheid blijven hebben om te vechten over rechts of links.</i> ‘In this way, children can vote for the things that will affect them. While the parents continue to have the opportunity to fight over the right from the left.’
That ... if construction (dat/als of dat/wanneer-constructie)	Conditional clause that immediately follows the main sentence while it could have been better at the end (e.g., <i>I think that if he behaves badly, he should go</i> , rather than: <i>I think that he should go if he behaves badly</i>).	Uit een onderzoek in 2012 is gebleken <i>dat wanneer</i> ouders meekregen dat wat hun kinderen allemaal deden op het internet ze een totaal nieuwe kant van hun kinderen leerde kennen. ‘A study in 2012 showed <i>that when</i> parents learned what their children were doing on the internet, they learned a

		completely new side to their children.'
<hr/>		
Stylistic lapse (other)***		
<hr/>		
Inappropriate words (ongepaste woorden)	Using words unbecoming of the text genre (e.g., swear words or words with unintended connotations).	Dus als de 16-jarige gaan stemmen wordt dit hele land <u>verkloot</u> door <u>imbicielen</u> die gaan stemmen. 'So when the 16-year-old starts to vote, this entire country is <i>screwed up by imbiciles</i> who start voting'
Spoken language (Gesproken taal)	Word use that would normally fit with spoken language, but not so well in written language.	Of ze <u>nou</u> een telefoon of een laptop van iemand anders lenen of dat ze met een een of andere app komen, hun ouders komen er toch niet achter. 'Whether they borrow someone else's phone or laptop or come up with some kind of app, their parents won't find out.'
Incorrect use of words (Woorden incorrect gebruikt)	Words used incorrectly, e.g., using a word where another word was intended.	Kinderen kunnen tegenwoordig ALLES online vinden zonder ook maar enige verantwoording af te staan. Ik vind dit <u>een abominabel geval</u> . 'Nowadays, children can find EVERYTHING online without giving any responsibility. I think this is an abominable case.'
Wrong use of commas (Kommafouten)	Using commas in the wrong places, in violation of formal guidelines for comma use.	Het stemmen is een groot ding in <u>Nederland, we</u> zijn vrij om onze eigen mening te uiten en de meeste doen dit dan ook met een kruisje in een stemhokje. 'Voting is a big thing in the Netherlands, we are free to express our own opinion and most of them do so with a cross in a voting booth.'

Mistakes in questionmarks (Vraagtekenfouten)	Ending a question with a full stop, or ending a declarative sentence with a question mark.	Wat voor partij en mensen zouden voor jou de beste maatregelen <u>treffen</u> . 'What kind of party and people would take the best measures for you.'
Colon and semicolon mistakes (Fouten tegen dubbele punt en puntkomma)	Wrong use of colon and semicolon (e.g., mixing them up).	Mijn stelling <u>is ouders</u> hoeven niet altijd mee te kijken met het internetgedrag van hun kinderen. 'My statement is parents do not always have to watch the internet behavior of their children.'
Abundant article use (Overvloedig gebruik van lidwoorden)	Using a definite or indefinite article while the context does not call for it, or if it does not allow for an article.	Ten eerste hebben kinderen ook recht op privacy, ook als het hun ouders zijn. <u>De</u> kinderen praten soms over dingen die zij niet met hun ouders willen bespreken en als <u>de</u> ouders deze geheimen of acties uitvinden kan dat de relatie tussen de ouders en het kind verslechteren. 'Firstly, children also have the right to privacy, even if they are their parents. The children sometimes talk about things that they do not want to discuss with their parents and if the parents find out these secrets or actions, the relationship between the parents and the child may deteriorate.'
Left out preposition/conjunction (Weggelaten voorzetsel/voegwoord)	Preposition or conjunction that has been left out in places where this should not be the case.	Dit heeft zijn <u>voordelen, het</u> opzoeken van informatie via internet voor bijvoorbeeld een spreekbeurt. 'This has its advantages, looking up information via internet for, for example, a speech.'
Erroneous preposition/conjunction	Using a wrong preposition or conjunction.	Als ouders mee zouden kunnen kijken zouden ze hiervan op de hoogte zijn en hun kind kunnen

(Foutief voorzetsel/voegwoord)		behoeden <u>van</u> eventuele gevaarlijke of ondoordachte situaties. ‘If parents could watch, they would be aware <u>for</u> this and could protect their child from any dangerous or ill-considered situations.’
Double prepositions (Dubbel voorzetsel)	Using two prepositions rather than one.	<u>Over</u> sommige onderwerpen praat een kind liever niet <u>mee</u> met zijn ouders maar wel op het internet. ‘On some topics a child prefers not to talk to his parents but to the internet.’
Derailed sentences (Ontspoorde zin)	A sentence that is made up of two (or more) conflicting sentence patterns, resulting in an ungrammatical sentence.	Stel je eens voor alles wat je op het internet doet, bijv. WhatsApp met iemand die je leuk vindt <u>dat jouw ouders dit allemaal lezen en alles over je weten</u> . ‘Imagine everything you do on the internet, for example texting with someone you like that your parents will read all this and know everything about you.’

* Note Unless indicated otherwise

** We aimed to keep translations as close to the original example as possible, although not all errors in Dutch will be considered errors or lapses in English.

*** Based on a bottom-up analysis that was informed by results from Van de Gein (2012) and Steenbakkers (2018)

Appendix B - Collinearity and linearity check

Collinearity

An issue in the analysis of this type of data is the collinearity of the explanatory variables. If the explanatory variables correlate too highly, the estimates of the individual parameters (and their standard errors) in the a regression analysis might be inflated (e.g. Mason & Perreault, 1991). Therefore, we checked the relations between the standard stylistic lapses predictors (see Table 4) as well as the other stylistic lapses predictors (see Table 5) in two ways.

First of all we estimated the correlations between the predictor variables. In both datasets the average correlation proved to be low (see Table A1). Furthermore, the standard deviation appears to be relative small, so there are not that many extremely low or extremely high correlations. The lowest correlation in both datasets concerns the relation between the dummy-variables for statement and position. The highest correlation for standard syntactic lapses is between the relative amount of erroneous non-finite clauses and the relative amount of agreement errors ($r = .25$),. In the data set with stylistic variables between two variables indicating the use of punctuation (relative amount of colon and semicolon mistakes and relative amount of mistakes in question marks, $r = .33$).

Table A1. Indices for collinearity of the explanatory variables (Sd: Standard deviation)

	Minimum	Maximum	Mean	Sd
<i>Standard stylistic lapses (relative measures; see Table 4)</i>				
Correlation	-.40	.25	-.01	.11
VIF	1.00	1.03	1.00	.01
<i>Stylistic lapses (relative measures; see Table 5)</i>				
Correlation	-.40	.33	.03	.13
VIF	1.00	1.18	1.02	.04

The second indicator of collinearity is a measure for variance inflation factors (VIF). This measure is based on r^2 , which quantifies the explained variance in a dependent variable by a set of explanatory variables. In a VIF-analysis each variable serves as a dependent variable which is explained by all other variables in the data set. A standard interpretation of VIF is that values over 4 indicate that multicollinearity might be an issue, while value over 10 indicate that collinearity is likely an issue. The estimates show that the explanatory variables In both the syntactic as well as in the stylistic data set can hardly be explained by the other variables. Both analyses show that collinearity is not likely to have inflated the parameter estimates in the present analysis.

Linearity

An assumption in regression analysis is that the relation between dependent and explanatory variables is linear. We have checked the linearity of the relation between text quality (dependent variable) and the explanatory variables by the formulation of a model which allows for nonlinearity. This has checked for each explanatory variable in a model in which next to a linear effect also quadratic and cubic effects of the explanatory variable were added to the model. The results are summarized in Table A2.

Table A2. Significance of cubic and quadratic effects of explanatory variables on text quality

	Effect			
	Quadratic		Cubic	
	Nonsignificant	Significant	Nonsignificant	Significant
Syntactic var.	82% (9)	18% (2)	82% (9)	18% (2)
Stylistic var.	83% (10)	17% (2)	83% (10)	17% (2)
Total	83% (19)	17% (4)	83% (19)	17% (4)

Results show that for 83% of the standard lapses variables (9 of 11) the cubic or the quadratic effect did not reach significance. So, for only 18% of the variables (2 out of 11) in this set the cubic and or the quadratic component reached significance (at $p < .05$). In the other lapses variable set the results do not differ much: 83% of the variables did not reach a quadratic effects or a cubic effect.

Of course in this analysis the significance of a lot of parameters (23 quadratic + 23 cubic = 46) is tested. Therefore, some significant effects are to be expected. A binomial test learns that the probability to observe 8 out of 46 test is over 17%. Consequently we conclude that the number of significant nonlinear relations is not likely to be a major issue in this data set.

Appendix C – Pearson correlation matrix among the relative measures of the stylistic lapses

	Pleonasm	Tautology	Contaminations	Ambiguous/vague referencing	Norm violations in referencing	Agreement errors	Dat/als-construction (that ... if)	Incorrect contractions	Incorrect inversion	Erroneous non-finite clauses	Detached phrases	Ongepast taalgebruik	Spoken language	Incorrect use of words	Wrong use of commas	Mistakes in question marks	Colon and semicolon mistakes	Abundant article use	Wrong article use	Left out p/c	Erroneous p/c	Double prepositions	Derailed sentences	
Pleonasm	1																							
Tautology	-.03	1																						
Contaminations	.04	-.05	1																					
Ambiguous/vague referencing	-.01	-.02	-.01	1																				
Norm violations in referencing	-.08	-.12	.18*	.12	1																			
Agreement errors	-.01	-.10	.06	.04	.12*	1																		
Dat/als-construction (that ... if)	.18*	.02	-.15	-.08	-.11	.07	1																	
Incorrect contractions	-.01	.24**	-.18*	-.07	.02	.12	.08	1																
Incorrect inversion	.07	-.08	-.00	-.13	-.08	.01	.23*	.08	1															
Erroneous non-finite clauses	-.03	-.04	.01	-.04	.05	.25**	.05	-.06	-.05	1														
Detached phrases	-.15	-.01	-.01	.24**	.08	.08	.17	.04	.09	-.07	1													
Incorrect use of words	-.06	-.07	.15	.02	.01	-.06	-.02	-.11	.01	-.03	-.03	1												
Spoken language	-.05	-.02	.14	-.13	.04	-.01	-.09	-.13	.07	-.08	-.05	.13	1											
Inappropriate words	.15	.09	.27**	.09	.07	-.01	-.10	.06	-.06	-.06	-.07	.32**	.15	1										
Wrong use of commas	.11	.02	-.11	.13	-.03	-.07	.25**	.02	-.04	.02	.06	-.02	-.13	.09	1									
Mistakes in question marks	-.08	-.04	-.05	.02	.04	.18*	.12	.08	-.04	.03	.24**	-.08	-.07	-.09	.29**	1								
Colon and semicolon mistakes	-.05	.22*	-.12	-.09	-.12	.06	.10	.17	-.12	.01	.00	-.08	-.07	.01	.33**	.33**	1							
Abundant article use	-.03	.03	.02	-.02	.16	.01	-.01	-.05	-.10	-.01	-.05	-.09	-.03	.09	-.06	-.00	-.06	1						
Wrong article use	.01	-.09	.17	.07	.04	-.01	-.09	.00	.01	-.04	-.04	.06	-.03	.20*	.08	.11	.01	-.01	1					
Left out p/c	-.06	.08	.04	.00	-.02	.01	-.08	.02	-.04	-.05	-.13	.08	.11	.24**	.11	.08	.01	.08	.03	1				
Erroneous p/c	-.01	.11	.35**	.03	.03	-.05	-.04	-.11	-.17	-.10	.08	.19*	.05	.22*	.20*	.14	.19*	-.04	.13	.10	1			
Double prepositions	-.07	.01	-.04	.15	-.06	-.06	-.08	-.01	-.01	-.03	-.08	-.06	-.02	-.05	-.05	-.01	.08	-.13	.16	-.11	-.03	1		
Derailed sentences	-.08	.05	.07	.01	.01	.15	.08	.11	.13	-.08	.25**	-.05	.16	.01	.04	.13	-.01	.06	-.02	-.06	.07	-.03	1	

*Note. * = significant at the 0.05 level, ** = significant at the 0.01 level