

# AN EXAMPLE OF AN L2-WRITING STRATEGY: NOVSKEV

FRANCIS KOX & HUUB VAN DEN BERGH

*Utrecht University*

## Abstract

In this study the effect of a writing strategy for L2-writing is tested experimentally. 10<sup>th</sup> grade students ( $N = 67$ ) took part in a writing experiment in writing German as a second language. In the experimental condition students were taught to pay attention to pre-writing activities (think and organize), find the right words in German, and post-writing activities (evaluate and revise).

Students in both the experimental and control condition completed two texts as pre- and post-test. All texts were rated holistically as well as with an analytic scoring scheme. Results show that in the experimental condition students wrote (on average) better texts than in the control condition, when rated holistically. For the analytic ratings improvement was statistically significant for the categories Content, Vocabulary and Conventions. For other categories (Syntax, Grammar (Verb and Case), Spelling and Punctuation) no effect of the experimental program could be shown.

Keywords: writing, writing strategies, foreign language (German), secondary education

1

*Kox, F. & van den Bergh, H. (2018). An example of an L2-writing strategy: NOVSKEV. Contribution to a special issue in honor of Gert Rijlaarsdam Making Connections: Studies of Language and Literature Education. L1-Educational Studies in Language and Literature, 18, pp. 1-12. <https://doi.org/10.17239/L1ESLL-2018.18.03.06>*

Corresponding author: Huub van den Bergh, Utrecht University, Trans 10, 3512 JK Utrecht, The Netherlands. Email: [h.vandenbergh@uu.nl](mailto:h.vandenbergh@uu.nl)

© 2018 International Association for Research in L1-Education.

## 1. INTRODUCTION

In the last decades teaching students to write has been an active area of research. Not only have many studies have been carried out in many countries (e.g. Fidalgo, Torrance, Rijlaarsdam, van den Bergh, & Álvarez (2015) in Spain, Rijlaarsdam et al (2011) in the Netherlands, or MacArthur, Graham & Fitzgerald (2008) in the US, to name a few), but also a number of meta-analyses have been carried out as well (e.g. Graham & Perrin, 2007; Graham, McKeown, Kihara & Harris, 2012; Koster, Tribushinina, De Jong & Van den Bergh, 2015). One of the most effective ways of teaching students to write are so-called writing strategies. Writing strategies are relatively fixed steps a writer carries out successively during writing. In many studies the different steps are shortened as a kind of mnemonic device for instance *POW*: **P**ick an idea, **O**rganize notes and **W**rite a complete paragraph (Harris, Graham & Mason, 2002) or *DODO*: **D**enken (think), **O**rden (organize), **D**oen (do) and **O**verlezen (read the whole text; Bouwer, Koster & Van den Bergh, 2018). Research shows that the learning of writing strategies is not only effective in primary education but in secondary education as well. For instance, in the meta-analysis of Graham and Perin (2007) and Graham et al. (2012) strategy instruction proved to be effective for students in grades 4 through 12 (with effect sizes varying from .92 to 1.02). In many writing programs the learning and application of strategies is supplemented by teacher or student modelling of the specific steps, and or by teacher and peer feedback as input for revision of the first version (e.g. Koster, Bouwer & Van den Bergh, 2016; Elving & van den Bergh, 2017).

The teaching of writing to students in their own language can be very effective, as the reported effect sizes are moderate or large (Graham & Perin, 2007; Graham et al., 2012; Koster et al., 2015). Writing is difficult because writers have to meet many requirements, while at the same time; they have to think of the audience, structure the text in a logical way, formulate their sentences well, think of what they want to say, etc. Altogether, writing is likely to cause cognitive overload, or to put it differently, writers experience difficulties in managing the writing processes (e.g. Beauvois, Olive & Passerault, 2011; Breetvelt, Van den Bergh & Rijlaarsdam, 1994). Writing strategies give a more or less fixed way to carry out a writing task: e.g. first you have think about what to write, after that you have to organize the ideas you just have come up with, and so on. So, by paying attention to successive steps in the writing process, attention is directed at each step separately. Consequently, cognitive overload is less likely.

L1-writing strategies have proven to be effective for different grades in different countries. However, students learning an L2 experience a manifold of difficulties when they write in the L2 as compared to writing in the mother tongue; above and beyond the requirements of writing texts in L1 as they have to find the words and use grammatical constructions which are not as automatized as in their L1 (Van Weijen, 2002; Tillema, 2012). These activities require cognitive resources which are already limited. Thus, it is not surprising that English texts of Dutch 14 year olds are on

average far below the average of their Dutch texts (ES = 2.0; Tillema, Van den Bergh, Rijlaarsdam & Sanders, 2013). If students experience 'problems' while writing in an L2, they are likely to switch to their L1 in order to circumvent working memory overload (Van Weijen, Van den Bergh, Rijlaarsdam & Sanders, 2019). Students writing in an L2 might benefit from writing strategies in the same way as if writing in their L1. Hedges (1988) gives a theoretical overview of L2-writing strategies. She argues that subdividing the L2-writing process of students into five phases resulted in a progress in text quality. Writers have to start with a so-called composing phase, in which they think what they want to write. In the second phase, 'communication', writers think about their public: they have to think on how the intended public can be reached, which information the (intended) public already possesses, and which words, phrases and style fit the (intended) public. In the third phase language plays a crucial role. This phase is very relevant for L2-writers as they have to look up words in a dictionary or search for synonyms. In the fourth phase students write their texts. In the fifth and final phase students re-read and revise their texts. Especially peer feedback is assumed to be useful for rewriting texts. Unfortunately, Hedge does not give criteria for successful peer feedback. Furthermore, it is stressed that the students should write functional texts; texts that are related to the students' lives and situations they encounter in their daily activities.

Of course, writing strategies are only part of effective education. According to the all-ready mentioned meta-analysis, explicit instruction, prewriting activities, (peer) feedback and revision activities are already ingredients of effective writing programs (in L1). Modelling, in which the teacher or peers carry out (part of) the writing assignment is mentioned less frequently (e.g. Couzijn, 1995; Rijlaarsdam et al., 2005), but is assumed to be an effective ingredient for learning to write; students do not need to write themselves but all attention is directed at learning from the writing model.

In this study a general writing strategy which proved to be effective for writing in L1 (Bouwer & Koster, 2016; Elving & Van den Bergh, 2017) was transformed to a writing strategy for writing in German as an L2. In this strategy students are required to carry out pre-writing activities (Hedge, 1988; Bouwer & Koster, 2016; Elving & Van den Bergh, 2017). A special activity, during prewriting is called 'Germanise', which indicates a phase in which writers can consult a dictionary (Dutch-German), a thesaurus or the website [www.uitmundend.de](http://www.uitmundend.de). After finishing the first draft each student received feedback from two classmates on his texts, after which each student rewrote his text.

## 2. METHOD

### 2.1 Participants

In this study 91 10th grade students of higher general vocational education (havo) participated. All students chose German as subject, and must complete a writing

exam at B1-level of the CEFR in the subsequent year (Fasoglio, Beeker, De Jong, Keuning & Van Til, 2014). At the start of the year students were randomly allocated into three classes. In each class German was taught by a different teacher. Students who were absent during one or both writing tests did not participate in the study. From the 91 potential participants, 78 students took the pre-test and 71 took the post-test, making a total of 67 students for both tests. In one class the experimental writing program was delivered to the students, while the other two classes serve as control groups.

The experimental and control condition were comparable with respect to the term results (experimental condition:  $M = 6.96$ ;  $sd = .88$ ; control condition:  $M = 6.64$ ;  $SD = 1.23$ ;  $t(65) = 1.17$ ;  $p = .25$ ). Furthermore, the percentage of boys (experimental condition 38% of 26; control condition 39% of 41;  $\chi^2(1) = .02$ ;  $p = .96$ ).

## 2.2 Design

All students wrote two persuasive texts, one as a pre-test and one as post-test. For the pre-test students had to write a persuasive letter to the head of the German section of the school (who doesn't want to go to Berlin on the next school trip and students have persuade him to change his mind). The post-test consisted of a persuasive letter as well. In this letter students had to convince the head of the school that the building should be painted in a more colourful way; the grey colour painted during the school's construction is not acceptable any more.

With the pre-test the initial level of writing was assessed, and with the post-test, which was taken after the experimental writing program the final level of writing was determined. All texts were written on a computer at school during German class. For both texts students had to finish their texts within one lesson (of 45 minutes). Students could make use of many aids (such as dictionaries, or the website [www.uit-muntend.de](http://www.uit-muntend.de)). However, translation programs were not allowed). The experimental program was delivered only in one class; writing was taught in both control classes for the same amount of time (i.e. 4 lessons of 45 minutes). The post-test was a school wide test for all students.

## 2.3 Class time

Based on the studies by Bouwer and Koster (2016), Braaksma (2002) and Hedge (1988) we developed a general writing strategy for students of German in the 10<sup>th</sup> grade. In this writing strategy the writing process is divided into seven different steps, in which students' attention is subsequently drawn to prewriting activities, characteristics of German language and revision activities. The acronym of this writing strategy is NOVSKEV, which stands for **N**achdenken (think), **O**rdnen (Organize), **V**erdeutschen (Germanise), **S**chreiben (write), **K**orrekturlesen (re-read), **E**valuieren (evaluate) and **V**erbessern (revise). Note that these steps have shown to be effective in previous studies in L1 (Bouwer & Koster, 2016; Graham & Perrin, 2007; Graham et

al, 2012; Koster et al, 2015) with the exception of *Verdeutschen* (Germanise) which is advocated by Hedge (1988)

NOVSKEV rests on four pedagogical pillars, which have proven to be effective in L1-writing: students working together (e.g. Van Steendam, Rijlaarsdam, Van den Bergh & Sercu, 2014)) and peer feedback (e.g. Graham & Perrin, 2007; Elving & Van den Bergh, 2017), modelling (e.g. Braaksma, Rijlaarsdam, Van den Bergh & Hout Wolters, 2004), direct instruction (e.g. Koster et al 2015).

The teacher of the experimental class introduced the writing strategy NOVSKEV in the first lesson. In several lessons she modelled the different phases of NOVSKEV. In the first lesson she explained the ideas behind writing strategies and handed out a paper with the explanation of the seven steps of NOVSKEV. Afterwards she modelled the first three steps **N**achdenken (think), **O**rdnen (Organize) and **V**erdeutschen (Germanise)) with a writing assignment which resembled the writing task with which students had to practice later on in this lesson. During modelling the teacher showed that writing is not easy, and that some ideas that popped up during brainstorming (**N**achdenken) were disregarded later on during **O**rganizing the ideas. Finally, **V**erdeutschung, can be challenging: the teacher showed that the choice for specific words was indeed difficult. In fact, the teacher showed in front of the class how she first gathers information (from long term memory), organized that information, followed by a phase in which she reflected on how to write in German and to look for translations in a dictionary. During modelling students were allowed to talk and deliberate, and ask questions. Once the three steps were modelled and all questions were asked the students could start with their assignment.

The second lesson, of 50 minutes, followed logically from the first, as the students had to write a text from their organized (and translated) ideas. Every student wrote his text and interaction between students was minimized during this phase. Ten minutes before the end of the second lesson all students had finished their first version of the letter. In the remaining ten minutes a classmate read the text and gave comments on the text.

The third lesson started with a modelling phase in which the teacher revised her text (which was shown on a so-called smartboard) while thinking aloud. After that students read their own text very precisely and revised their text when they deemed necessary or based on the comments of their classmate. They had to pay attention to lower order (grammar, spelling) as well as to higher order aspects (organization, content, style) of their text. When they finished their text, the text was read by a classmate who gave feedback on the content, organization and grammar.

In the fourth and final lesson students had to rewrite their letter a last time. In Table 1 a summary of teacher and student activities per writing lesson is presented.

In the control condition, students practiced with the same assignment, and were stimulated to reread and revise their text in a second lesson. They were not taught any of the elements of the writing strategy as outlined above. The lessons consisted for the most part in learning writing by doing which is business as usual in the Netherlands (compare, Henkens, 2010).

Table 1. A summary of the learning of the writing strategy NOVSKEV.

Lesson	Step in strategy	Activity student		Activity teacher	
		Working together	Peer Feedback	Modelling	Frontal instruction
1	Think (Nachdenken)	✓		✓	✓
1	Organize (Ordnen)	✓		✓	✓
1	Germanize (Verdeutschen)			✓	✓
2	Write (Schreiben)				
3	Correct (Kurekturlesen)	✓	✓	✓	✓
3	Evaluate (Evaluieren)	✓	✓	✓	✓
4	Revise (Verbessern)				✓

#### 2.4 Rating of text quality

Texts from the pre- and post-test were rated with the same rating scheme, which has been used at this school for several years (see, Table 2). According to this rating scheme students start with the maximum score of 12 and a number of different errors lead to subtraction of a number of points, e.g. using the wrong word in this context leads to a subtraction of .6 points each time a word is used wrongly, with a maximum of three points (see Table 2). For content students could earn 3 points if the goal of their letter was clear, and there were convincing arguments for their point of view, which was well organized, and the tone of the letter was appropriate to the social context. If a letter failed on either of these criteria one or more points were subtracted. One of the advantages of this rating model is that it gives students feedback on several aspects of the quality of their texts. The number of deduction points per type of error is grounded in the tacit knowledge of the teachers at this school.

It has, however, been argued that a holistic score might be more valid (see, Van den Bergh, De Maeyer, Van Weijen & Tillema, 2012). Therefore, all texts were rated in a more holistic way as well. To give both raters some support during rating they were given an average text with reasons why this was an average text with an arbitrary rating of 100. Each essay was compared with the example. If the essay in question was superior to the example text, the rating was higher, and if the essay was poorer in quality than the example, the rating was lower.

The correlation between both raters, teachers of the school at which the experiment was carried out, proved to be 0.70. Therefore, the reliability of these holistic ratings is estimated as .82, which is quite high for the rating of essays (compare, Wesdorp, 1981).

Table 2. Analytic rating scheme used at the school where the experiment was carried out (see also Appendix).

Type of error	Max. deduction	Total Number of points: 12	
		Deduction per error	Number
Vocabulary: <i>wrong word in this context</i>	3 points	-0,6	
Syntax: <i>"Klammerstellung" after conjunction</i>	1 point	-0,5	
Grammar: <i>Error in verb</i>	2 points	-0,3	
Grammar: Grammatical case		-0,2	
Spelling: <i>Typo's</i>	1 point	-0,1	
Content: <i>see text</i>	3 points	N.A.	
Punctuation: <i>Use of capital and comma's</i>	1 point	More than 10 errors -1	
Conventions: <i>e.g. s mall letter after salutation and comma after ending</i>	2 points	-1	
Number of points and final mark:			/ 1.2 =

The correlation between (the average) holistic ratings and (the average) analytic ratings was  $r = 0.77$ , which is relatively high (corrected for unreliability the correlation even equals .94). Therefore, the holistic and analytic ratings result in (approximately) the same ranking of texts, and we cannot conclude that one of the types of ratings is less valid than the other.

### 3. RESULTS

In Table 3 the average ratings for both pre- and post-test per condition are presented. This table shows that the average holistic rating in the experimental condition increases from 95.7 ( $SD = 17.7$ ) to 117.5 ( $SD = 16.9$ ). In the control group the average rating increases from 91.6 ( $SD = 17.2$ ) to 100.2 ( $SD = 17.1$ ). For the other rating categories (except Content), we needed to use subtraction points (see Table 1). Therefore, the lower the scores, the better the text.

Table 3. Average ratings per measurement and condition (standard deviation between brackets; N: Number of observations).

	Experimental Condition (N = 26)		Control Condition (N = 46)	
	Pre-test	Post-test	Pre-test	Post-test
Global quality	95.7 (.17)	117.5 (.17)	91.6 (.17)	100.2 (.17)
Content	1.00 (.66)	2.58 (.37)	.67 (.54)	.99 (.59)
Vocabulary <sup>1</sup>	.95 (.68)	.37 (.51)	1.27 (.86)	1.02 (.88)
Syntax: <i>Klammerstellung</i> <sup>1</sup>	.08 (.23)	.08 (.18)	.17 (.26)	.04 (.08)
Grammar (verb) <sup>1</sup>	.30 (.36)	.13 (.19)	.35 (.47)	.21 (.24)
Grammatical case <sup>1</sup>	.35 (.32)	.15 (.23)	.66 (.47)	.26 (.23)
Conventions <sup>1</sup>	.73 (.77)	.08 (.27)	.76 (.68)	.51 (.64)
Spelling errors <sup>1</sup>	.13 (.23)	.06 (.09)	.11 (.14)	.05 (.07)
Punctuation <sup>1</sup>	.10 (.15)	.09 (.16)	.10 (.14)	.09 (.19)

<sup>1</sup> The lower the better (see Table 2)

To test whether the differences between conditions and measurements are significant analysis of variance for repeated measurements was used. In this case the differences between measurement occasions is hard to interpret as students completed different writing assignments at both occasions.<sup>1</sup> Therefore, differences in means cannot be interpreted unequivocally, as they might be due to the progress of students, or to the differences between both assignments. The interaction between condition and measurement occasion is the effect of interest, because this interaction tells us that the difference between pre-test and post-test depends on the condition.

Results show that the (average) rating of the global quality depends on the interaction of condition and measurement occasion; in the post-test students in the experimental condition (on average) outperform students in the control condition (see also Table 3).

For analytic ratings an interaction effect of Measurement Occasion \* Condition was found for Content, Vocabulary, Syntax and Conventions. That is, on the post-test (average scores for) Content, Vocabulary and Conventions was better in the experimental condition as compared to (the averages in) the control condition. Students in the control condition made more Syntax errors on the pre-test, while on the post-test the number of errors did not differ significantly.

<sup>1</sup> Students wrote texts on different topics on pre- and posttest. If students were to write on the same topic on both occasions the post-test scores would have been inflated due to the fact that students had already written a text on that topic for the pre-test. This design, although usual in writing research, results in uninterpretable main effects of measurement occasion, as this effect consists of learning gain and/or differences in difficulty between assignments. The effect of the intervention is, however, tested by means of the interaction effect between measurement occasion and condition. Therefore, experimental effects are not confounded with assignment.

Table 4. Testing statistics (*F*) and significance (*p*) for effects of Measurement Occasion, Conditions and the interaction (Occ \* Condition; *df*: degrees of freedom).

	Occasion [ <i>df</i> (1, 65)]		Condition [ <i>df</i> (1, 65)]		Occ * Condition [ <i>df</i> (1, 65)]	
	<i>F</i>	<i>p</i>	<i>F</i>	<i>p</i>	<i>F</i>	<i>p</i>
Global quality	26.31	<.01	10.41	<.001	4.13	.046
Content	.39	.54	.13	.71	18.75	<.01
Vocabulary	10.71	<.01	10.83	.01	4.69	.03
Syntax	2.55	.12	.17	.69	4.98	.03
Grammar (verb)	9.77	.01	.94	.34	.12	.73
Grammatical case	32.32	<.01	10.41	<.01	3.23	.08
Conventions	20.11	<.01	3.65	.06	4.19	.04
Spelling errors	7.04	.01	.50	.48	.11	.74
Punctuation	.38	.53	.14	.71	.01	.98

#### 4. DISCUSSION

In this study we have explored the effects of a writing strategy for L2-writing. Based on the literature and on examples of effective writing strategies, a strategy was developed in which pre- and post-writing activities were specified. Special attention was paid to L2-difficulties of word finding (and use in a foreign language) as well as grammatical issues; writers had to pay attention to these aspects in particular.

Results of the experiment show that this strategy, which draws writer's attention to pre- and post-writing activities, is successful when general text quality is rated. Also an effect of the writing strategy could be shown for specific analytic categories: Content, Vocabulary, Syntax and Conventions. 10<sup>th</sup> grade writers appear to produce better texts when the writing strategy is used. It seems that for higher-order dimensions of text quality an effect of NOVSKEV could be shown, whereas for lower-order dimensions (punctuation, grammar (verb) grammatical case, spelling and punctuation) differences between conditions did not appear to be significant. With hindsight this might have been expected, as the writing strategy focusses explicitly on higher order aspects of texts (for instance, think on what you want to say (Nachdenken), organise your ideas (Orden), and students' comments mainly pertained to clarifying) and much less on lower order aspects. Maybe students should have been instructed explicitly to pay attention to the lower order aspects of their writing in the revision phase.

This study can be characterized as an exploratory investigation into the effectiveness of L2-writing strategies. It shows that L2-writing strategies can be effective, but it is still unclear whether the results can be generalized over teachers (is this strategy also effective if it is delivered by other teachers?), over languages (does the effectiveness depend on specific aspects of the L2), or over writing tasks. The latter is especially relevant as the correlation between texts of different writing prompts have shown to be low in both L1 and L2 (e.g. Bergh et al., 2012). Therefore, a criterion

which consists of multiple writing tasks is preferable (see for instance, Bouwer, Koster & Van den Bergh, 2017).

The presented writing strategy, NOVSKEV, as a whole has shown to increase the quality of texts written in German by Dutch Students. We did not study the effect of the different components of this strategy.

Finally, the agreement between holistic ratings and analytic ratings is remarkable. Most research on differences between rating procedures stem from L1-research. It may be that insights from L2-research do not generalize to other languages that easily. A similar conclusion was drawn by Van Weijen who showed that a) the quality of L2-texts did not differ as much as the quality of L1-texts of the same writer, and b) that the writing process of L2-writers did not differ that much over different writing assignments as the writing process of the same writers in their L1.

#### REFERENCES

- Beauvais, C., Olive, T., & Passerault, J. M. (2011). Why are some texts good and others not? Relationship between text quality and management of the writing processes. *Journal of Educational Psychology*, 103(2), 415-428. <https://doi.org/10.1037/a0022545>
- Braaksma, M. (2002). *Observational learning in argumentative writing*. Amsterdam, The Netherlands: UvA.
- Bergh, H. van den, Maeyer, S. de, Weijen, D. van & Tillema, M (2012). Generalizability of text quality scores. In E. van Steendam, M. Tillema, G. Rijlaarsdam & H. van den Bergh (Eds), *Measuring writing* (p.23-32). Brill, The Netherlands: Leiden.
- Berninger, V. W., Vaughan, K., Abbott, R. D., Begay, K., Coleman, K. B., Curtin, G., ... & Graham, S. (2002). Teaching spelling and composition alone and together: Implications for the simple view of writing. *Journal of Educational Psychology*, 94(2), 291-304. <https://doi.org/10.1037/0022-0663.94.2.291>
- Bouwer, I.R. & Koster, M. (2016). *Bringing writing research into the classroom. The effectiveness of Tekster, a newly developed writing program for elementary students* (UIL-OTS/ICO-series). Utrecht, The Netherlands: Universiteit Utrecht.
- Bouwer, R. Koster, M. & Bergh, H. van den (2017). Improving writing skills in upper elementary grades: An intervention study. *Journal of Educational Psychology*, 108(1), 58-71. <http://dx.doi.org/10.1037/edu0000206>.
- Braaksma, M. Rijlaarsdam, G., Bergh, H. van den & Hout Wolters, B.A.M. (2004). Observational learning and its effects on the orchestration of the writing process. *Cognition & Instruction*, 22(1), 1-36. [https://doi.org/10.1207/s1532690Xci2201\\_1](https://doi.org/10.1207/s1532690Xci2201_1)
- Breetvelt, I., Bergh, H. van den, & Rijlaarsdam, G. (1994). Relations of writing processes and text quality: When and how. *Cognition and Instruction*, 12(2), 103-123. [https://doi.org/10.1207/s1532690Xci1202\\_2](https://doi.org/10.1207/s1532690Xci1202_2)
- Couzijn, M. (1995). Observation of writing and reading activities; effects on learning and transfer (dissertation). Amsterdam, The Netherlands: University of Amsterdam
- Elving, K. & Bergh, H. van den (2017). Doen we weer booster? Het effect van een digitale en interactieve schrijfcursus op tekstkwaliteit in havo 4 [Are we doing BOOSTER again? The effect of a digital interactive writing course]. *Pedagogische Studiën*, 94, 330 - 347.
- Fasoglio, D. Beeker, A. Jong, K. de, Keuning, J. & Til A. van, (2014). *ERK-niveau schrijfvaardigheid Engels, Duits en Frans Onderzoek naar het bereikte niveau aan het eind van havo en vwo [CEF levels of writing skill in English, German and French. A study to the level of writing skills at the end of general higher vocational education and pre-university education]*. Enschede, The Netherlands: SLO.
- Flower, L., & Hayes, J. R. (1980). (1980b). The dynamics of composing: Making plans and juggling constraints. In: L.W. Gregg & E.R. Steinberg (Eds.), *Cognitive processes in writing* (pp. 31-50). Hillsdale, NJ: Lawrence Erlbaum.

- Fidalgo, R., Torrance, M., Rijlaarsdam, G., van den Bergh, H., & Álvarez, M. L. (2015). Strategy-focused writing instruction: Just observing and reflecting on a model benefits 6th grade students. *Contemporary Educational Psychology, 41*, 37-50. <https://doi.org/10.1016/j.cedpsych.2014.11.004>
- Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology, 99*(3), 445-476. <https://doi.org/10.1037/0022-0663.99.3.445>
- Graham, S., McKeown, D., Kiuahara, S., & Harris, K. R. (2012). A meta-analysis of writing instruction for students in the elementary grades. *Journal of Educational Psychology, 104*(4), 879-896. <https://doi.org/10.1037/a0029185>
- Harris, K. R., Graham, S., & Mason, L. (2002). POW plus TREE equals powerful opinion essays. *Teaching Exceptional Children, 34*(5), 74-77. <https://doi.org/10.1177/004005990203400513>
- Hedges, T. (1988). *Writing*. Oxford, UK: Oxford University Press.
- Henkens, L.S.J.M. (2010). *Education in writing of texts [Het onderwijs in het schrijven van teksten]* (pp. 1-58). Utrecht, The Netherlands: Inspectorate of Education.
- Koster, M. P., Tribushinina, E., De Jong, P., & van den Bergh, H. H. (2015). Teaching children to write: A meta-analysis of writing intervention research. *Journal of Writing Research, 7*(2), 299-324. <https://doi.org/10.17239/jowr-2015.07.02.2>
- MacArthur, C. A., Graham, S., & Fitzgerald, J. (Eds.). (2008). *Handbook of writing research*. New York: Guilford Press.
- Rijlaarsdam, G., Braaksma, M., Couzijn, M., Janssen, T., Raedts, M., Van Steendam, E., Toorenaar, A. & Van den Bergh, H. (2008). Observation of peers in learning to write: Practice and research. *Journal of Writing Research, 1*(1), 53-83. <https://doi.org/10.17239/jowr-2008.01.01.3>
- Rijlaarsdam, G., Bergh, H. van den, Couzijn, M., Janssen, T., Braaksma, M., Tillema, M., Steendam, E. van, Raedts, M. (2011). *Writing*. In Graham, S., Bus, A. Major, S. & Swanson, L. (Eds.), *Application of educational psychology to learning and teaching*. APA Handbook Volume 3 (pp. 189-228). Washington, DC: American Psychological Association. 10.1037/13275-001.
- Steendam, E. van, Rijlaarsdam, G. C. W., Van den Bergh, H. H., & Sercu, L. (2014). The mediating effect of instruction on pair composition in L2 revision and writing. *Instructional Science, 42*(6), 905-927. <https://doi.org/10.1007/s11251-014-9318-5>
- Tillema, M. (2012). *Writing in first and second language. Empirical studies on text quality and writing processes*. Utrecht, The Netherlands: LOT.
- Tillema, Bergh, H. van den, Rijlaarsdam, G., Sanders, T. (2013). Quantifying the quality difference between L1 and L2 essays: a rating procedure with bilingual raters and L1 and L2 benchmark essays. *Language Testing, 30*(1), 71-97. <https://doi.org/10.1177/0265532212442647>
- Weijen, D. Van (2008). *Writing processes, text and task effects: Empirical studies in first and second language writing*. Utrecht, The Netherlands: LOT.
- Weijen, D. van, Bergh, H. van den, Rijlaarsdam, G. & Sanders, T. (2009). L1 use during L2 writing: An empirical study of a complex phenomenon. *Second Language Writing, 18*(4), 235-250. <https://doi.org/10.1016/j.jslw.2009.06.003>
- Wesdorp, H. (1981). *Evaluatietechnieken voor het moedertaalonderwijs*. [Evaluation techniques for mother tongue education]. 's Gravenhage, The Netherlands: Staatsuitgeverij.

## APPENDIX

*Description of the categories in the analytic rating scheme.*

Category	Description	Example
Vocabulary	When a word is used in the wrong context, and the meaning of the sentence becomes unclear.	<b>Die Lehrlinge</b> vom Isala wollen mehr Farbe in der Schule.
Syntax: Satzklammer	The finite verb in a subordinate clause is not last.	Wir möchten gerne eine Reise nach Berlin, weil das unsere Deutschkenntnisse <b>kann verbessern</b> .
Grammar: verbs	The verb is not conjugated properly.	<b>Ich weiß</b> e nicht, ob das geht.
Grammar: case	The case is not properly declined.	Ich habe <b>eine gute Zweck</b> gewählt, <b>es</b> heißt KiKa.
Spelling	Typo's	Es gibt da einen großen <b>Unterscheid</b> .
Punctuation	Commas and uppercase letters are ignored.	Der gute <b>zweck</b> will gerne <b>geld</b> sammeln für <b>kinder</b> .
Conventions	If no comma follows after the salutation and if a comma follows after the final salutation.	<b>Sehr geehrte Frau Schnitzler</b> Mit freundlichem <b>Gruß</b> ,