Developing students' writing in History: Effects of a teacher-designed domain-specific writing instruction

Jannet van Drie, Johan van Driel & Daphne van Weijen

University of Amsterdam | The Netherlands

Abstract: Writing in history places high demands on students and is a skill that requires explicit instruction. Therefore, teachers need to be able to teach this in an effective way. In this study, the writing-instruction was designed by a teacher, instead of researchers, as part of a professional development program in the Netherlands. The lessons combined writing and historical reasoning instruction, based on principles of effective writing instruction, including strategy-instruction, modeling, prewriting, and peer-interaction. The effects of these lessons were investigated in a small-scale pilot study, which consisted of a pre-test post-test quasi-experimental design, in which eighty-nine 11th grade students participated (39 in the treatment condition and 50 in the comparison condition). Dependent measures included text quality, writing process measures, students' knowledge of writing and their self-efficacy. Students in the treatment condition wrote longer and higher quality texts, spent more time writing, paused more while writing and their knowledge of writing was higher at post-test than for students in the comparison condition. No effects were found for self-efficacy. Furthermore, significant correlations were found between text quality and writing process measures, but not for knowledge of writing and self-efficacy. Overall, the effectiveness of this teacher-designed intervention seemed satisfactory, as it resulted in greater knowledge of writing and better-quality writing in his history classes.

Keywords: historical reasoning, literacy, writing instruction, writing processes, secondary education.

Within the field of writing research, more attention has been paid in recent years to writing in the disciplines (see for example Rijlaarsdam & Braaksma, 2015). Initially the focus was on effects of writing activities on content knowledge; the so-called writing-to-learn approach (Applebee, 1984; Bangert-Drowns, Hurley & Wilkinson, 2004; Emig, 1977). Recently, more attention has been paid to learning to write discipline-specific text genres and accompanying lines of reasoning, in order to increase students' literacy skills. Shanahan and Shanahan (2008) pointed out that literacy develops from generalizable basic skills to more specialized disciplinary literacy skills in middle and high school. As each discipline has its' own forms of discourse and rhetorical demands, students should be taught how to read and write various text-genres from different disciplines (Shanahan & Shanahan, 2008; van Drie, Janssen, & Groenendijk, 2017). Also, in the Netherlands, the context of this study, more attention is given to reading and writing in different school subjects. With respect to educational policy, the referential levels for language (Expert Group Learning Trajectories, 2009), for example, explicitly state that attention to language, reading and writing is not only a matter for the language subjects, but for all school subjects. The Language Oriented Content Teaching approach provides teachers with teaching practices that integrate content and language teaching (den Elzen, Oorschot, Visser et al., 2018; Hajer & Meestringa, 2020;).

For the discipline of history, the focus of this study, various researchers stressed the importance of discipline-specific writing instruction (e.g., De La Paz & Felton, 2010; Moje, 2008; Monte-Sano, 2010; Nokes & De La Paz, 2018; van Drie, Braaksma, & van Boxtel, 2015). History is predominantly a literate discipline that includes the analysis of (often) written historical sources and the (re-)construction of interpretations of the past in written form. Writing is one of the means in which students' reasoning about the past can be expressed and developed, for example through source-based writing or short-answer questions. Monte-Sano (2010) argued that writing is essential to learn the substantive and procedural forms of knowledge of history. Writing in history places high cognitive demands on students (Coffin, 2006), as they have to combine historical content knowledge and historical reasoning skills, with knowledge of rhetorical demands of the text. This entails metacognitive knowledge of what constitutes a good text and which writing strategies can be used to write such a text (McCutchen, 2011; Schoonen, van Gelderen, de Glopper et al., 2003).

Within the Dutch history curriculum, the most common form of writing is answering short answer questions as it is used as a learning activity to develop knowledge as well as in testing. In addition, the central examinations consist of open-ended short answer questions, contrary to, for example, document-based essay writing as used in Anglo-Saxon countries. Teachers' experience is that students have difficulty in expressing their knowledge and ways of reasoning in written text (den Elzen et al., 2018). However, history teachers, in the Netherlands

and abroad, often do not seem to pay explicit attention to writing in their lessons and, if they do, they restrict themselves to requirements of the writing product, such as structure and layout and hardly any attention is paid to the writing processes involved, such as generating ideas, organizing, and revising, or to genre-characteristics, goal and audience awareness or language use (De Oliveira 2011; Gillespie, Graham, Kiuhara, & Hebert, 2014; McCarthy Young & Leinhardt, 1998; Mottart, van Brabant, & van de Ven, 2009). In addition, Monte-Sano and Allen (2019) showed in their study on writing practices in U.S. history classrooms, that a variety of approaches to writing exist among (novice) teachers. For many (history) teachers, paying attention to writing seems time-consuming as it distracts them from teaching content and they consider teaching writing a part of the language arts curriculum (Moje, 2008). This might be problematic as there are indications that students hardly transfer their knowledge of writing learned in the language arts to other content areas (Mottart et al., 2009).

It thus seems challenging to stimulate sustainable change in teacher practices, to ensure that they will include writing instruction in their history lessons. Clarke and Hollingsworth (2002) indicated that teacher change or professional growth is, amongst other variables, influenced by professional experimentation and experiencing positive student learning outcomes, which in turn influence teachers' beliefs and attitudes. In writing research, interventions are usually designed by researchers rather than teachers, which might be a potential barrier for teachers' long-term implementation of successful interventions (Borko, 2004; Koster, Bouwer, & van den Bergh, 2017; Koster, Tribushinina, de Jong, & van den Bergh, 2015). A possible solution might be that teachers design the writing-interventions themselves, based on design principles (van Drie et al., 2017) and investigate the effects themselves so that they might see evidence of their students' positive learning firsthand. Therefore, in this pilot study, we investigated the effects of domain-specific writing instruction lessons, designed by a history teacher as part of a professional development program (the effects of this program on teacher beliefs is reported in Van Drie et al., 2017). In addition to the effects of the intervention on text quality, we investigated effects on students' writing processes, meta-knowledge of writing and students' self-efficacy.

1.   Theoretical Framework

1.1   Effective writing in history

Earlier research on writing indicated that, in order to write well, students must have two main types of knowledge: domain-specific knowledge on the one hand and strategic knowledge on how to solve the task at hand on the other (Beaufort, 2004). For domain-specific knowledge, Beaufort proposed a conceptual model of five types of knowledge which expert writers rely on while writing: discourse-

community knowledge, subject-matter knowledge, rhetorical knowledge, genre knowledge and writing-process knowledge (Beaufort, 2004, p. 141). Strategic knowledge is related to how these different types of knowledge can be applied while writing (Beaufort, 2004; Hayes & Flower, 1980). Subject-matter knowledge, or content knowledge, is important as a writer must know what to write about. It is therefore not surprising that writers with more relevant content knowledge tend to write better texts (Braaksma, van den Bergh & Rijlaarsdam, 2018; McCutchen, 1986).

For writing in History, we would make a distinction between historical content knowledge and historical reasoning competencies. Historical reasoning can be defined as: 'attempts to reach justifiable conclusions about processes of continuity and change, causes and consequences, and/or differences and similarities between historical phenomena or periods' (van Boxtel & van Drie, 2018, p. 151). It includes aspects such as contextualization and argumentation based on critical analysis of historical sources. Writing is an important means for expressing historical reasoning (van Drie et al., 2015). Research indicated several difficulties students encounter, such as developing arguments and the use of evidence (e.g., McCarthy Young & Leinhardt, 1998; Nokes & De La Paz, 2018), the contextualization of historical events and persons (Sendur, van Drie, & van Boxtel, 2021) or incorporating specialized historical terms, the substantive concepts and second order concepts (i.e., cause, change) (Stoel, van Drie, & van Boxtel, 2017). However, proficient writing in history requires more than being able to reason historically, as shown in a study by Stoel et al. (2017). They investigated the effects of an explicit teaching intervention on causal reasoning in a randomized controlled trial with 95 pre-university students (grade 11). Although they found positive effects on students' knowledge of causal reasoning strategies and second-order concepts, no effect was found on the quality of students' written products. This suggests that applying one's understanding of historical reasoning strategies in a written text requires an additional step. It also requires knowledge of the text structure of historical explanations in history and rhetorical knowledge of how these discipline specific ideas can be best represented in text (Langer, 1992).

With respect to strategic knowledge of writing processes, Hayes and Flower (1980) originally discerned three main processes involved in writing: 1) planning, which includes generating ideas, organizing and setting goals; 2) translating domain-specific knowledge into language under the control of the writing plan; and 3) reviewing the text, in which the text is evaluated by means of the goals; this involves reading the text and possibly correcting or editing it. These three processes are supervised by the monitor and operate upon two kinds of information; (1) the task environment, which consists of the writing assignment (topic, goal, instructions) and the text produced so far, and (2) the writers' long-term memory, which contains topic knowledge, knowledge of the audience, and linguistic knowledge about text plans and grammar rules. According to Hayes & Flower,

expert writers are able to construct a more sophisticated representation of their goals and develop and modify this representation throughout the writing process, which enables them to revise more extensively, and to evaluate their texts on the basis of their goals. Moreover, the process of generating goals and organizing knowledge to satisfy these goals can lead to the discovery and generation of new ideas.

Writers thus also need knowledge of characteristics of a good text and of the processes of how to achieve such a text. In his revised model of writing Hayes (1996) proposed that writers should have knowledge of schemata of different text genres; their structure, their components and knowledge of typical linguistic markers to relate these components (see also Hayes, 2012). This knowledge can subsequently facilitate planning and revision processes (McCutchen, 2011). Within the context of school-history, several text-genres are used, such as biographical recount, historical account, explaining and arguing genres, each with their own structure and linguistic demands (Coffin, 2006). These linguistic demands are intertwined with the historical reasoning required (see also Kellogg, 2008).

Furthermore, it is essential that writers learn how and when to apply both their content knowledge and knowledge of writing in order to be able to write good quality texts (Beaufort, 2004; Kellogg, 2008). Knowledge of text genres and writing processes is referred to as meta-cognitive knowledge of writing (Bouwer & Koster, 2016; Harris, Graham, Brindle & Sandmel, 2009; Schoonen & de Glopper, 1996). This metacognitive knowledge is found to be positively related to writing performance; successful writers have more declarative, procedural and conditional knowledge about writing than less successful writers (Bouwer & Koster, 2016; Klein & Kirkpatrick, 2010; Schoonen & de Glopper, 1996). However, students' knowledge about writing is usually restricted to, and mostly directed at, the written product and hardly at writing processes (Bouwer & Koster, 2016; Schoonen & de Glopper, 1996; Schoonen et al., 2003). There are indications that writing instruction might improve students' metacognitive knowledge of writing (Bouwer & Koster, 2016; Klein & Kirkpatrick 2010; van Drie, Janssen, & Groenendijk, 2018). Thus far, less is known about how this works within the disciplines. In addition, little or no research has addressed measuring students' online writing processes in history classes. This is unfortunate, as data on writing processes might help us understand how knowledge about writing, writing processes and text quality are linked in writing in history (van Drie et al., 2015). Keystroke logging tools such as Inputlog (Leijten & van Waes, 2013) can provide valuable insights into writers' online writing behaviour, without possible intrusions caused by other methods such as concurrent think-aloud protocols. Therefore, in this this study we included measurements on knowledge of writing and writing processes.

1.2 Writing instruction in History

From the above it can be concluded that writing in general, and in history, is a complex and demanding task (cf. Rijlaarsdam, Braaksma, Couzijn et al., 2005) and explicit and domain-specific writing instruction might be needed to develop students disciplinary writing abilities. Meta-studies provide indications of effective writing instruction (e.g., Graham & Perin, 2007; Graham & Harris, 2018a; 2018b). In this study we focus on writing strategy instruction, studying text models, prewriting activities and collaboration during writing, as these function as design principles for the intervention. Some of these approaches have also been studied within the field of history. Thus far, the best investigated approach for teaching domain-specific writing is strategy instruction. An example of strategy instruction is the self-regulated strategy development model (SRSD: Harris & Graham, 1996; see also Graham, Harris, MacArthur & Schwartz, 1998;), which considers writing to be a purposeful activity and uses an expert-novice apprenticeship model to lead students to independent use of writing heuristics. SRSD includes five stages for instruction: develop background knowledge, describe it, model it, support it, and independent performance (Harris & Graham, 1996). De La Paz and colleagues used this model for teaching domain-specific writing, adapting it for teaching reasoning with historical documents, and found positive effects of this approach on students' writing in several studies (e.g., De La Paz, 2005; De La Paz & Felton, 2010).

Studying text models was applied in history class in a study by Van Drie et al. (2015). This product-oriented approach to teaching writing, involves analyzing text examples to establish criteria for a good text (Hillocks, 1986). In this study, students were provided with three examples of an introduction, a body and a conclusion. By comparing the text-examples in small groups students determined criteria for good writing. The outcomes were subsequently discussed within the whole class. The idea behind this approach is that by discovering criteria themselves and making them explicit, students will apply these criteria more readily in their own texts. In their meta-analysis Graham and Perin (2007) found positive but small effects of this approach.

The use of prewriting activities and of collaboration have been studied in history less often to date. One example of a prewriting activity is the construction of a scheme to select and order information before writing. Van Drie, van Boxtel, Jaspers, & Kanselaar (2005) compared the effects of the construction of three different schematic formats (i.e., argumentative diagram, matrix and list) on the quality of the texts produced. Participants, 11th grade students, worked in pairs on a historical inquiry task in a computer-based learning environment. Although no effects were found on text quality, the schematic format did direct students' historical reasoning in their chat conversations during task execution. Although this study included collaborative writing as a component, its effect could not be determined, as students in all conditions collaborated with a peer.

Despite these and other studies conducted in the field of writing in history, more knowledge is needed about effective writing instruction in history (Nokes & De La Paz, 2018). In this study we will investigate the effects of a teacher-designed intervention, based on design principles, on text quality, writing processes, students' knowledge of writing and students' self-efficacy.

1.3 Aims and research question

This study aims to contribute to the existing body of research on writing in history in two different ways. First, although domain-specific writing instruction designed by researchers has been shown to be effective, its implementation by teachers in classroom settings might be hampered by teachers' lack of involvement in the design process (Koster et al., 2015). Earlier studies have indicated that it is important that teachers participate in the development of intervention studies (Borko, 2004; Koster et al., 2017). Therefore, we want to investigate whether writing instruction lessons designed by a teacher on the basis of design principles would results in higher quality texts as well. Second, as little is known about students' writing processes in history classes, their knowledge of writing and self-efficacy, we included these as dependent measures and also explored the relations between them.

The main research question guiding this study is: What are the effects of domain-specific writing instruction, designed by a history teacher, on the quality of students' writing texts in history class, their writing processes, knowledge of writing, and self-efficacy?

To investigate this question, we conducted a study in which a teacher designed lessons for 11th grade history students, in the context of a professional development program on domain-specific writing instruction. The effects of these lessons including a domain-specific writing instruction (treatment group) and the existing approach (comparison group) were compared on several outcome measures in a quasi-experimental pre- and post-test design. This made it possible for the teacher to gain insight into the possible effect of the redesigned lessons, on the quality of students' texts, their writing processes, knowledge of writing, and writing self-efficacy.

Based on the existing literature, we predict that students in the treatment group will produce higher quality texts and have more knowledge of writing than students in the comparison group and will also show an increase in self-efficacy. Students in the treatment group might be more likely to plan and reflect on their actions while writing and are likely to engage more with the sources than students in the comparison group. So, with regard to writing processes, we predict that students in the treatment group will pause more often and longer on average while writing than writers in the comparison group. Furthermore, we will explore whether there are

positive relations between text quality scores and students' writing processes, knowledge of writing and their self-efficacy.

2. Method

2.1 Participants

One history teacher (who is also the second author[1]) and his four classes ($N = 98$) participated in this study. The teacher holds a master's degree in history and had, at the time of the study, 12 years of teaching experience. He took part in a Professional Development (PD)-program aimed at integrating writing instruction in social sciences classes (the first author was one of the trainers). This PD involved six meetings, each of three and half hours, and was implemented during one schoolyear (see Van Drie et al. (2017) for an extended description of this program). The students were 16-17 years old, enrolled in 11th grade upper secondary education, and had no prior knowledge of writing in history, as it not explicitly part of the regular history curriculum in the Netherlands. Students did have content knowledge about the topics of writing for the pre and post writing task, as well as for the writing task of the intervention, as the topics were chosen in relation to the curriculum being currently taught. Also, students had experience with analyzing historical sources. Data were gathered in two cohorts in two consecutive school years, with two classes in each cohort. Within a cohort the two classes were randomly assigned to conditions. Students who did not gave permission for participating, who missed more than one of the intervention lessons, or who missed all the post measurements, were removed from the dataset, resulting in $N = 89$ (Ntreatment = 39; Ncomparison = 50). The two conditions did not differ significantly in their scores on the pre-writing task, $t(82) = .721$, $p = .47$; mean (*SD*): comparison group: 18.61 (4.24); treatment group: 17.94 (4.18).

2.2 Intervention

Design principles. The program of the PD focused on five design principles for effective writing instruction which the participating teachers could use to design a domain-specific writing intervention for one of their classes. The design-principles were selected by the trainers involved in the PD and derived from a meta-analysis of effective writing instruction by Graham and Perin (2007). These design principles were: 1) strategy-instruction using modelling, 2) studying text-examples, 3) using prewriting-activities, 4) peer-interaction and -review, and 5) using authentic tasks. In the first PD training session these design principles were explained and

---

[1] At the time when the study was conducted, he only had the role of teacher. He was only asked to act as co-author well after completion of the data collection.

illustrated with examples. A considerable part of the PD was used for (re-)designing lessons, including providing feedback on these lesson designs by the trainers and the other participants, during which the design principles and their implementation in different contexts were discussed. Teachers were free to choose which of the five design principles they wanted to use, and how to implement and adapt them to their specific teaching context. The teacher reported in this study started from lessons that he normally taught and redesigned them based on the design-principles mentioned above.

Writing task. The lessons centered around a writing task, which asked students to write an argumentative letter in response to a historian who wrote a book in which he claimed that the Dutch were partly responsible for the death of the Dutch Jews. Only a few Dutch citizens were willing to provide shelter for Jews, despite the fact that information was available at that time about the fate of Jews concentration camps. This latter point has been criticized by several Dutch historians. The students received a set of seventeen short historical sources, from which they could derive arguments in favor of, or against, this claim. The historical sources were 5 to 10 sentences each and included primary sources such as diary fragments and secondary sources. The text had to be about 1½ pages A4 in length in 12 point times roman font and was written on the computer during a single lesson.

The task could be considered an authentic history task (design-principle 5), as it requires students to critically examine a historical interpretation and argue, based on historical sources, whether they think this interpretation is valid. Furthermore, it is set in a realistic context, as there was quite some discussion in the media about this book and the author received letters from people agreeing or disagreeing with his interpretation. From a writing perspective, the task could be considered authentic as it has a clear goal and audience.

Lessons in the treatment group. The teacher designed a series of five lessons around the writing task that included instruction on writing in history for which the design-principles mentioned above were used (see Appendix A). In the first lesson the writing task was introduced. The teacher explained the point of view of the historian, the debate about this interpretation and the importance for students to develop their own interpretation based on historical sources. The task was also made relevant for the students by highlighting the importance of writing and arguing for the upcoming central examinations. Furthermore, background knowledge about the situation of the Jews in the Netherlands during the Second World War was also provided. In addition, the teacher provided instructions on argumentation in history, such as the use of arguments for and against, weighing arguments, and the use of contextual information, after which students completed some short assignments. Next, they worked in peers on assignments related to

background information on the topic and making assignments related to understanding the substantive concepts and on constructing a historical context. Lesson 2 focused on the role of source-analysis in history and on providing evidence for a specific historical interpretation. Students received a step-by-step guide on how to analyze sources, based on the three heuristics identified by Wineburg (1991). This included questions such as, who is the author (sourcing), taking into account the time the source was written (contextualization), what is the author's claim, what kind of evidence is used, and comparing the source to other sources (corroboration). After a short period of instruction by the teacher students analyzed the sources in dyads, using a worksheet in Excel. In lesson 3 the students received instruction on argumentative text structure, in which text-examples were studied. In addition, the teacher modelled writing an argumentative letter and students then worked on their own writing plan. Lesson 4 was spent on writing the text, based on their writing plan and using the historical background and the analysis of the sources. Finally, in lesson 5 students gave feedback on another student's text, using a rubric, and students then revised their own text based on this feedback.

Lessons in the comparison group. As mentioned before, the original lessons that the teacher had previously taught, were used as a comparison in this case. The writing task was introduced in the same way as for the treatment group. Also, the information on the historical context was the same for both groups, but this comparison group did not receive instruction on writing in history. In the first lesson the writing task was introduced and background knowledge on the topic at hand was developed, using the same information and assignments as the treatment group. The only difference was that students in this group did not receive instruction on argumentation in history. In the next lesson they wrote the text. So, compared to the treatment group, students received two lessons; lesson 1 (without instruction on argumentation) and lesson 4 (see Appendix A). Later that school year these students also received instruction on writing in history in a similar way as the treatment group, but with a different historical content.

2.3 Instruments and data

Text quality. To measure improvement in the quality of students' writing a pre- and post-writing task were administered. Both times students wrote a text similar to the one used in the intervention, an argumentative letter, although with different content as pre- and post-tasks. To ensure that students had sufficient content-knowledge, topics were chosen that were part of the curriculum at the time the test was administered. Students had one lesson (50 minutes) to complete these tasks. The length of the texts was 300 words and some historical documents were included (containing both arguments for and against the topic). The prewriting task was on

imperialism in the Dutch Indies, the Atjeh-War and, in particular, the role of the governor-general van Heutsz (1904-1909). Students were asked to write an argumentative letter to a newspaper in Amsterdam in response to recent discussions about whether the statue of van Heutsz in Amsterdam should be removed or not. The topic of the post-writing task was the Cold War. Students were asked to write an argumentative letter to a history magazine in which they responded to the statement 'The Soviet-Union is responsible for the Cold War'.

The pre- and post-texts were analyzed using the rubric that was originally used in the PD-program (van Drie et al., 2017; see also Appendix B). The rubric contained three criteria, each with three sub-criteria: (1) Genre-specific writing quality, with the sub-criteria Introduction, Body, Conclusion; (2) General writing quality, with the sub-criteria Audience orientation, Coherence, Language use and spelling; and (3) Domain-specific quality, with the sub-criteria Use of domain-specific concepts, Content, Use of sources. For each subcategory four levels were described, on a scale of 1 to 4, instead of the six levels in the original version of the rubric.

Two coders (the first author and a research assistant) coded and discussed several texts in a training-phase, after which they each calculated a sample of 48 texts (about 18% of all texts), equally spread over conditions and measurement points (pre-intervention, intervention, post-intervention). The interrater agreement for this blind coding session was determined using Cohens' Kappa and was found to be moderate to substantial (Mean Kappa: .586; range: .50 to .71; Viera & Garret, 2005). Correlations between raters' scores were high ($r = .74$ to .87). Because measuring text quality is an inherently subjective process, all texts were independently scored by the two coders, and correlations between them over all texts were high as well ($r = .68$ to .83). Subsequently, the mean of the two raters' scores per text at pre- and post-test were used for further analyses. Here we report on the pre- and post-test only.

Cronbach's alpha was calculated for the three-sub criteria of each category and turned out to be satisfactory to good for all three criteria for pre-test as well as post-test (range $= .67$ to .90). Subsequently, sum scores for each of the criteria at pre- and post-test were calculated. In addition, as Cronbach's alpha turned out to be satisfactory ($= .84$ and .93, respectively),-an overall text quality score for the pre-test texts and one for post-test texts was calculated as well.

Writing processes. Inputlog (Leijten & van Waes, 2013) was used to measure students' online writing in Word while they wrote the pre- and post-task texts. Inputlog registers every keystroke a writer uses and every pause or revision (s)he makes while writing, which means the tool generates very detailed process data, which can be analysed in multiple ways (see also Vandermeulen, Leijten, & Van Waes, 2020). As a result, researchers must choose which of the many available variables are relevant to include in the analyses for their specific study (Leijten &

Van Waes, 2013; Van Waes, Leijten & van Weijen, 2009). In line with earlier research, (e.g., van Drie, Groenendijk, Braaksma, & Janssen, 2016; Vandermeulen, 2020), we chose to include the total number of words produced during writing, the total number of words in the final text, and writing time in the analysis. In addition, we also chose to include pause time as a variable, in two ways: the number of pauses of a certain length and total pause time.

When including pauses in the analysis, it is important to choose a threshold for pause length, which indicates the minimum length of pauses that will be included in further analyses. The chosen pause threshold can range from very short (500ms or shorter) to much longer pauses (e.g. 1, 2, 5 or 10 seconds), depending on the focus of the study. Pause times during writing are generally considered indicative of cognitive effort and pause length seems to increase as text units become larger (Leijten & van Waes, 2013). In other words, pauses between paragraphs tend to be longer than pauses between sentences, which in turn are usually larger than between-word or within-word pauses. This means that if you are mainly interested in studying cognitive processes rather than linguistic level text production then it seems suitable to focus on longer pause times (Wengelin, 2006; Van Waes, Leijten, Lindgren & Wengelin, 2015). The most common pause threshold in earlier studies seems to be 2 seconds, but a higher threshold is also possible. In this case, because we are mainly interested in more complex cognitive processes which play a role in writing-to-learn, such as planning, reading, idea generating and problem solving, we chose to focus on pauses of 10 seconds or more, in line with earlier research by Van Drie et al. (2016).

Knowledge of writing. To measure (changes in) students' knowledge about writing, a short writing task was used at pre- and post-test. This task was based on Schoonen and de Glopper (1996), which asked students to write a short text (email) to advise a friend how to write an argumentative text in history. The texts were segmented and analyzed based on the number and the content of students' recommendations. We checked whether a recommendation was a) related to the text-genre, b) domain-specific and c) whether it was related to the writing product or to writing processes. For example, the recommendation 'You should support your claim with arguments from historical sources' was coded as: genre-specific (argumentation), domain-specific (history) and writing product and the recommendation 'First, make a list of the information you want to write about' was coded as: not genre-specific, not domain-specific and writing process.

Interrater agreement was calculated between the first author and a research assistant on 32 texts (285 segments) equally divided over the conditions and pre- and post-test measurements (about 18% of all texts). Interrater agreement was high: Cohens' Kappa varied between .89 and .96.

Self-efficacy. Students' self-efficacy for writing in general and for writing an argumentative text was measured using a questionnaire developed by Braaksma et al. (2018). The questionnaire contained sixteen items for general writing self-efficacy and seven for writing argumentative texts. For each item students could give points on a scale from 0 to 100. Examples of two items for general writing were: 'I can use connectives within a paragraph and between paragraphs.' and 'In my text I can take my audience into account'. Examples for argumentative texts were: 'I can convince the reader of my point of view with my text' and 'I can provide correct evidence for my arguments in my text'. Cronbach's alpha, calculated for the different scales for the pre- and post-test was found to be good (ranging from .90 to .92).

2.4 Analyses

To identify possible differences between conditions we used repeated measures analyses, with the factor time (pre- versus post-test) as the within subjects variable and condition (treatment vs comparison group) as the between subjects factor. These analyses were conducted for text quality, writing processes, knowledge of writing and self-efficacy. Furthermore, to explore possible relations between variables included in the study, correlations (Spearmans' r_s) at post-test were calculated between text quality and writing processes, knowledge of writing and self-efficacy.

3. Results

3.1 Text quality

First, we will report on the effects of the intervention on text quality. Table 1 presents the scores of the texts at pre- and post-test, as well as the outcomes of the repeated measures analyses. The maximum score for each of the three categories was 12, the overall maximum score was 36. The analyses showed that students in the treatment condition scored significantly higher on overall text quality, as well as on all three criteria. Figure 1 shows the scores on overall text quality for both conditions.

Repeated measures analyses revealed a significant interaction effect between time and condition on overall text quality, $F(1, 80) = 94.85$; $p = .000$, a medium effect (partial $\eta^2 = .54$). Students in the treatment condition scored significantly higher on the post-test compared to the comparison condition, both in terms of overall text quality as well as the three sub-criteria General, Domain-specific and Genre-specific text quality (see Table 1). The average text quality for students in the treatment group increased significantly from pre- to post-task for all these criteria, while the scores for the comparison group remained more or less stable (see Figure 1).

Table 1: Mean scores, standard deviations and outcomes for repeated measures analyses for text quality

Variable	Measurement occasion	Treatment group (N = 34)	Comparison group (N = 48)	Measurement occasion effect	Condition effect	Interaction effect
		Mean (sd)	Mean (sd)	(df 1 = 1; df 2 = 80)		
General Text quality	Pre-test	5.44 (1.52)	5.90 (1.74)	$F = 77.09$**	$F = 27.98$**	$F = 93.57$**
	Post-test	9.53 (1.98)	5.70 (1.68)	$\eta_c^2 = .49$	$\eta_c^2 = .26$	$\eta_c^2 = .54$
Domain-specific text quality	Pre-test	5.87 (1.63)	5.86 (1.57)	$F = 27.56$** $\eta^2 = .26$	$F = 17.51$**	$F = 44.90$** $\eta^2 = .36$
	Post-test	8.44 (1.80)	5.55 (2.15)		$\eta_c^2 = .18$	
Genre-specific text quality	Pre-test	6.28 (1.59)	6.85 (1.84)	$F = 33.27$**	$F = 5.56$*	$F = 43.12$** $\eta^2 = .35$
	Post-test	8.85 (1.61)	6.69 (1.91)	$\eta_c^2 = .29$	$\eta_c^2 = .07$	
Overall Text quality	Pre-test	17.59 (3.99)	18.61 (4.24)	$F = 70.70$**	$F = 19.70$**	$F = 94.85$** $\eta^2 = .20$
	Post-test	26.82 (4.74)	17.94 (5.07)	$\eta_c^2 = .47$	$\eta_c^2 = .54$	

All effects were significant at * = $p < .05$; ** = $p < .001$;
Interpretation of partial eta squared (η_c^2) effect sizes: 0.2 = small, 0.5 = medium, 0.8 = large.
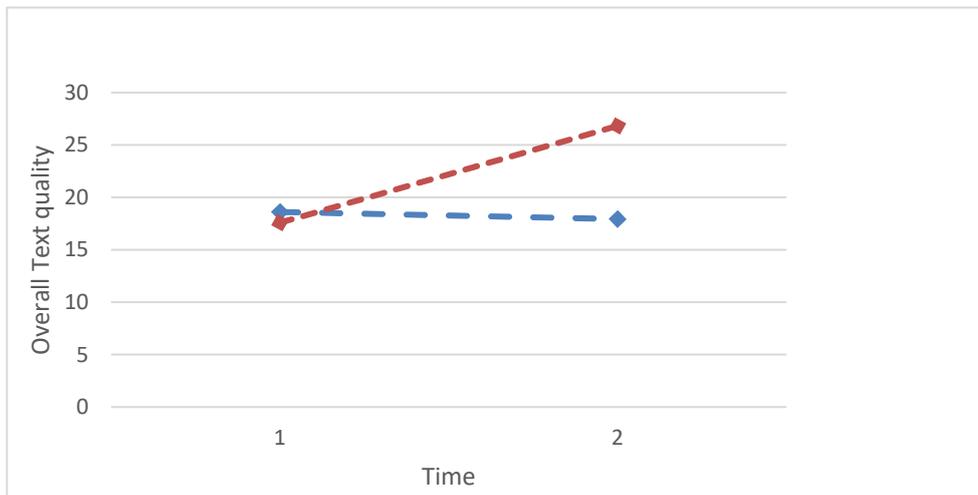

Figure 1. Overall text quality scores at pre- and post-test
Legend: Long dashes = comparison condition; Short dashes = treatment condition

3.2 Writing processes

Table 2 presents the outcomes of the analyses of students' writing processes. Repeated measures analyses showed a significant interaction effect for writing time, total word count, word count final text, number of pauses (10 seconds or longer) and pause time. Students in the treatment group spent significantly more time writing during the post-test task than students in the comparison group. In addition, they also wrote significantly more words during the post-test task and produced longer final texts (see Table 2). Finally, students in the treatment group took on average more pauses of 10 seconds or more and paused longer in general during the post-test task than students in the comparison group.

Table 2. Overview of outcomes for repeated measures analyses for writing process measures

Variable	Measurement occasion	Treatment group (N = 30) Mean (sd)	Comparison group (N = 41) Mean (sd)	Measurement occasion effect (df 1 = 1; df 2 = 69)	Condition effect	Interaction effect
Writing time (in seconds)	Pre-test	1193.49 (194.90)	1062.21 (269.78)	$F = 14.68**$ $\eta^2 = .17$	$F = 37.14**$ $\eta^2 = .35$	$F = 24.49**$ $\eta^2 = .26$
	Post-test	1230.47 (259.62)	771.68 (237.38)			
Total word count	Pre-test	397.13 (86.97)	397.66 (79.08)	n.s.	$F = 11.44**$ $\eta^2 = .14$	$F = 39.93**$ $\eta^2 = .37$
	Post-test	472.40 (123.19)	329.00 (108.06)			
Word count final text	Pre-test	308.90 (55.89)	325.29 (43.61)	n.s.	$F = 13.55**$ $\eta^2 = .16$	$F = 42.04**$ $\eta^2 = .38$
	Post-test	382.43 (91.64)	275.49 (64.77)			
Number of 10 sec or longer Pauses	Pre-test	23.23 (6.25)	20.61 (6.28)	$F = 25.11**$ $\eta^2 = .27$	$F = 29.81**$ $\eta^2 = .30$	$F = 18.97**$ $\eta^2 = .22$
	Post-test	22.60 (9.23)	11.56 (4.46)			
Pause time 10 sec or longer	Pre-test	781.18 (230.44)	790.24 (263.99)	n.s.	$F = 18.03**$ $\eta^2 = .21$	$F = 41.02**$ $\eta^2 = .37$
	Post-test	938.86 (300.18)	497.74 (234.45)			

Significance levels: * = $p < .05$; ** = $p < .001$; n.s. = non-significant

Interpretation of partial eta squared (η^2) effect sizes: 0.2 = small, 0.5 = medium, 0.8 = large.

3.3 Knowledge of writing

Table 3 presents the outcomes of the analyses of students' knowledge of writing. Positive significant interaction effects were found for average number of recommendations, number of genre-specific recommendations, domain-specific recommendations, and for product recommendations. For each of these measures, students in the treatment group gave significantly more recommendations on average at post-test than students in the comparison group. However, no effects were found for the number of process related recommendations.

Table 3. Overview of outcomes for repeated measures analyses for knowledge of writing

Variable	Measurement occasion	Treatment group ($N=33$)	Comparison group ($N=49$)	Measurement occasion effect	Condition effect	Interaction effect
		Mean (sd)	Mean (sd)	(df 1 = 1; df 2 = 80)		
Average no. of recommendations given per student	Pre-test	8.61 (3.79)	9.02 (4.15)	n.s.	$F = 5.15*$ $\eta^2 = .06$	$F = 18.21**$ $\eta^2 = .18$
	Post-test	9.64 (2.28)	6.31 (3.08)			
Genre-specific recommendations	Pre-test	4.03 (2.87)	4.16 (2.71)	n.s.	$F = 5.49*$ $\eta^2 = .06$	$F = 11.65**$ $\eta^2 = .13$
	Post-test	5.42 (2.11)	3.22 (2.16)			
Domain-specific recommendations	Pre-test	.70 (1.13)	.53 (.77)	n.s.	$F = 11.26**$ $\eta^2 = .12$	$F = 7.12**$ $\eta^2 = .08$
	Post-test	1.12 (1.22)	.20 (.76)			
Number of Product related recommendations	Pre-test	7.09 (3.20)	7.43 (4. 75)	n.s.	$F = 5.14*$ $\eta^2 = .06$	$F = 18.29**$ $\eta^2 = .19$
	Post-test	8.27 (3.48)	4.73 (3.23)			
Number of Process related recommendations	Pre-test	1.52 (2.15)	1.41 (1.29)	n.s.	n.s.	n.s.
	Post-test	1.36 (1.85)	1.43 (1.85)			

Significance levels: * = p < .05; ** = p < .001; n.s. = non-significant

Interpretation of partial eta squared (η^2) effect sizes: 0.2 = small, 0.5 = medium, 0.8 = large.

3.4 Self-efficacy

Table 4 presents the means and standard deviations for the self-efficacy questionnaire on general writing and argumentative writing at pre- and post-test. The repeated measures analyses revealed no differences between conditions for students' self-efficacy related to general writing or argumentative writing.

Table 4. Mean scores and standard deviations on self-efficacy at pre-test and post-test

Variable	Measurement occasion	Treatment group		Comparison group	
Self-efficacy		N	Mean (sd)	N	Mean (sd)
General writing	Pre-test	30	71.33 (11.28)	48	69.56 (10.63)
	Post-test		72.01 (8.78)		70.68 (11.87)
Argumentative writing	Pre-test	29	71.24 (9.43)	45	70.28 (13.74)
	Post-test		73.91 (8.33)		73.73 (11.00)

3.5 Relations between variables

Correlational analyses were carried out to determine whether the different variables included in the study were interrelated and whether these correlations differed between conditions. First, correlations were calculated to determine possible relations between text quality and writing process measures at post-test for both conditions. Results indicated that overall text quality correlated positively with the Inputlog measure total number of words produced for both groups, although the correlation was somewhat larger for the treatment group (see Table 5). In addition, overall text quality was positively correlated with the number of words in the final text for the treatment group but not with the other variables. By contrast, overall text quality was significantly correlated with total writing time for the comparison group, but not with the other variables.

Table 5. Significant correlations (Spearmans' r_s) between overall text quality and writing process variables per condition

	Overall text quality per condition	
Writing Process variables	Treatment Group ($N = 31$)	Comparison Group ($N = 45$)
Total no. of words produced	.44*	.30*
Words final text	.36*	n.s.
Writing time	n.s.	.30*
Number of pauses	n.s.	n.s.
Pause time	n.s.	n.s.

Significance levels: * = $p < .05$; ** = $p < .001$; n.s. = non-significant

Next, we calculated correlations between text quality and knowledge of writing per condition. The outcomes revealed no significant correlations between those variables. Lastly, we calculated correlations between text quality and self-efficacy. Results indicated that there were no significant correlations between text quality

and students' self-efficacy for general writing or between text quality and students' self-efficacy related to argumentative writing in the treatment and comparison groups.

4. Conclusion and Discussion

In order to write well in history, in addition to content knowledge and historical reasoning skills, students also need to have knowledge of characteristics of a good text and of writing processes. As a result, writing in history places high demands on students and therefore teachers need more knowledge about how to teach these skills in an effective way. Although research has provided us with several examples of effective domain-specific writing instruction, it is important that teachers design the interventions themselves, adapted to their specific teaching context and thereby gain insight into the effects of the intervention. This might contribute to changing teaching practices (Clarke & Hollingsworth, 2002; Borko, 2004; Koster et al, 2017). In this pilot study we investigated the effects of a teacher-designed domain-specific writing instruction for 11th grade history students on text quality, writing processes, knowledge of writing and self-efficacy.

With respect to text quality, we found, as expected, that students in the treatment group wrote significantly better texts at post-test, compared to the comparison group. This positive effect was found for overall quality, as well as for the three criteria general text quality, domain-specific quality and genre-specific text quality. With respect to writing processes we found that students in the treatment condition took more pauses of 10 seconds or more and paused longer in general at post-test than students in the comparison group, as was expected. In addition, these students spent more time writing and wrote longer texts. The fact that comparison group students wrote for less time, on average, on the post-test task than the pre-test task was a somewhat surprising finding. Shorter writing time could be a sign of task disengagement, because students were tired or bored or did not like the topic of the second task and thus felt less motivated to write a long text about it. This possible explanation is supported to some extent by the positive correlation between overall text quality and writing time we found for the comparison group; Students who took more time were likely to write better texts (see Table 5). However, the possible disengagement explanation is contradicted to some extent by the fact that students in the comparison group did not write texts of poorer quality at post-test than at pre-test. The average overall text quality score for the comparison group did not differ significantly between the two measurement occasions (see Table 1). If students were really disengaged, then we would have expected the quality of their texts also to be much poorer for the post-test task.

Regarding other measures, we found that students' knowledge of writing increased significantly in the treatment condition. Students not only gave more recommendations, but also more genre-specific, domain-specific (although

relatively few) and product-related recommendations. This suggests that the domain-specific writing instruction provided students with a better idea of what was expected of them when writing an argumentative text in history. Exploratory correlational analyses showed that text quality was positively related to writing processes, particularly for the treatment group, but not to students' knowledge of writing or their self-efficacy.

The positive outcomes of this small-scale intervention on text quality are in line with earlier studies that found positive effects of domain-specific writing instruction on students' writing in history (e.g., De La Paz & Felton, 2010; Monte-Sano, 2010; Nokes & De La Paz, 2018; van Drie et al., 2015). It also adds to these findings, as, in this case, the teacher designed the domain-specific writing instruction himself, based on a set of design principles. Despite the fact that only one teacher participated in this study, it is a promising finding that when a history teacher designs the intervention himself, it can also lead to increased text quality, knowledge of writing and more effective writing processes. Moreover, it shows that a history teacher might be able to integrate writing instruction in his own lessons in an effective manner, even though he is not specialized in writing instruction. This outcome is also supported by the results of a study on the effectiveness of the PD-program the teacher took part in, which revealed that although participation in the PD program did not appear to influence teachers' beliefs about writing, it did influence their classroom practices to some extent (van Drie et al., 2017). Furthermore, the participating teachers appeared to feel more ownership of their lessons because they were able to implement the design principles in their lessons in their own way instead of having to implement researcher-developed lessons. Therefore, design principles might form a bridge between research literature and teaching practices and are well suited to be used in professional development programs (see also Graham & Perin, 2007).

In this study we included measurements of text quality, as well as on writing processes, knowledge of writing and self-efficacy. Thus far, little research has been conducted on writing processes in history. The positive effects of the intervention on students' writing processes and the outcome that text quality was positively correlated to writing processes in the treatment condition, show that this is an interesting field for further research. The outcome that the intervention resulted in more knowledge of writing is in line with earlier research (Bouwer & Koster, 2016; Klein & Kirkpatrick, 2010). However, no positive outcomes were found with regard to knowledge of writing processes and, overall, relatively few process-related recommendations were made. This is in line with earlier findings showing that students seem to possess less knowledge about writing processes (Bouwer & Koster, 2016; Schoonen & de Glopper, 1996; Schoonen et al., 2003). Although the intervention also addressed some issues related to writing processes, the main focus was on the writing product, which might explain these findings. As the

teacher taught writing for the first time, a focus on writing processes might have been a step too far. Future research could investigate the effects of domain-specific writing instruction that explicitly includes writing processes. In this study no correlation was found between text quality and students' knowledge of writing, which contradicts earlier findings (Bouwer & Koster, 2016; Klein & Kirkpatrick, 2010). More research is needed on the mediating role of knowledge of writing for subject-specific writing in upper secondary education. With regard to students' self-efficacy no effects were found. Longer interventions, than this one of only five lessons, might be needed to improve students' self-efficacy in domain-specific writing. Future research could also include writing motivation, as it seems to correlate strongly with writing performance (Wright, Hodges, Enright, & Abbott, 2021).

A drawback of the design of this study is fact that the two groups differed in two respects. First, the distribution of students over the conditions was uneven. However, although the comparison group was larger than the treatment group, we still found positive effects of the intervention. Second, students in the comparison group received two lessons, whereas the treatment group received five lessons and it could be argued that the effects might be related to the teaching time involved instead of the domain-specific writing intervention. For that reason, we did not include the text written during the intervention in the analyses and focused on the post-test writing task instead. Within the specific context and purposes of this study, it seemed a fair choice to compare the redesigned lessons with the lessons the teacher normally taught, as this would increase the ecological validity of the study. The lessons in the comparison condition reflected the situation in terms of how the teacher had approached writing in his lessons without domain-specific writing instruction, as many other teachers in the Netherlands commonly do. From a practical perspective it was interesting to make this comparison, as it is a comparison between regular and new teaching practices, and it investigated the effects of the new approach using quantitative methods instead of drawing mainly on personal experiences. The resulting evidence of the positive effect of this approach on student outcomes may result in more sustainable changes in teaching practices (Clarke & Hollingsworth, 2002). Only one teacher participated which has consequences for the generalizability of the outcomes of this study. As this study took place in the context of teacher development it made sense to involve only one teacher and to investigate learning outcomes of the new approach compared to the old one. Furthermore, this was a pilot study, with only four classes and using more teachers might have resulted in possible teacher effects.

To conclude, this pilot study shows that a teacher-designed domain-specific writing instruction has the potential to affect positively students' text quality, their writing processes and their knowledge of writing. Still, more research is needed on the comparison of effects of different domain-specific writing approaches, to

provide teachers with various examples of how to include writing instruction in their history lessons.

Acknowledgements

References

Applebee, A. (1984). Writing and reasoning. *Review of Educational Research, 54*, 577–596. https://doi.org/10.3102/00346543054004577

Bangert-Drowns, R. L., Hurley, M. M., & Wilkinson, B. (2004). The effects of school-based writing-to-learn interventions on academic achievement: A meta-analysis. *Review of Educational Research, 74*, 29-58. https://doi.org/10.3102/00346543074001029

Beaufort, A. (2004). Developmental gains of a history major: A case for building a theory of disciplinary writing expertise. *Research in the Teaching of English, 39*(2), pp. 136-185. https://www.jstor.org/stable/40171658

Borko, H. (2004). Professional Development and Teacher Learning: Mapping the Terrain, *Educational Researcher, 33*(8), 3–15. https://doi.org/10.3102/0013189X033008003

Bouwer, R. & Koster, M. (2016). *Bringing writing research into the classroom. The effectiveness of Tekster, a newly developed writing program for elementary students.* Unpublished Dissertation, University of Utrecht. https://issuu.com/tekster6/docs/tekster_proefschrift-digi2

Braaksma, M., & Rijlaarsdam, G. (2015). Introduction to the special issue on Writing-to-learn studies. *Journal of Writing Research, 7*(1), 1-4. https://doi.org/10.17239/jowr-2015.07.01.01

Braaksma, M. A. H., Rijlaarsdam, G., & van den Bergh, H. (2018). Effects of hypertext writing and observational learning on content knowledge acquisition, self-efficacy, and text quality: Two experimental studies exploring aptitude treatment interactions. *Journal of Writing Research, 9*(3), 259-300. https://doi.org/10.17239/jowr-2018.09.03.02

Clarke, D. & Hollingsworth, H. (2002). Elaborating a model of teacher professional growth. *Teaching and Teacher Education, 18*, 947-967. https://doi.org/10.1016/S0742-051X(02)00053-7

Coffin, C. (2006). *Historical discourse: The language of time, cause and evaluation.* New York: Continuum.

De La Paz, S. (2005). Effects of historical reasoning instruction and writing strategy mastery in culturally and academically diverse middle school classrooms. *Journal of Educational Psychology, 97*(2), 139-156. https://doi.org/10.1037/0022-0663.97.2.139

De La Paz, S. & Felton, M. K. (2010). Reading and writing from multiple source documents in history: Effects of strategy instruction with low to average high school writers. *Contemporary Educational Psychology, 35*, 174-192. https://doi.org/10.1016/j.cedpsych.2010.03.001

den Elzen, M., Oorschot, F., Visser, A., van der Leeuw, B. & Meestringa, T. (2018*). Redeneren en formuleren bij mens- en maatschappijvakken. [Reasoning and formulating in the social science subjects].* Enschede, The Netherlands: SLO.

De Oliveira, L. C. (2011). *Knowing and writing school history. The language of students' expository writing and teachers' expectations.* Charlotte, NC: Information Age Publishing.

Emig, J. (1977). Writing as a mode of learning. *College Composition and Communication*, 28, 122-128. https://doi.org/10.2307/356095

Expert Group Learning Trajectories (2009). *Referentiekader taal en rekenen: De referentieniveaus. [Reference framework language and arithmetic: The referential levels].* Enschede, The Netherlands: Ministerie van OCW.

Gillespie, A., Graham, S., Kiuhara, S., & Hebert, M. (2014). High school teachers use of writing to support students' learning: A national survey. *Reading and Writing, 27*(6), 1043-1072. https://doi.org/10.1007/s11145-013-9494-8

Graham, S., & Harris, K. (2018a). Evidence-based writing practices: A meta-analysis of existing meta-analyses. In R. Fidalgo & T. Olive (Series Eds.) & R. Fidalgo, K.R. Harris, & M. Braaksma, (Vol Eds.), *Design principles for teaching effective writing*. Studies in Writing Series: Vol. 34 (pp. 13-38). Leiden, The Netherlands: Brill. https://doi.org/10.1163/9789004270480_003

Graham, S. & Harris, K. (2018b). An examination of the design principles underlying a self-regulated strategy development study. *Journal of Writing Research 10*(2). 139-187. https://doi.org/10.17239/jowr-2018.10.02.02

Graham, S., Harris, K., MacArthur, C. A., & Schwartz, S. S. (1998). Writing instruction. In B. Y. L. Wong (Ed.), *Learning about learning disabilities* (2nd ed., pp. 391–424). San Diego, CA: Academic Press. https://doi.org/10.1016/B978-0-12-762530-0.50017-5

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

Hajer, M. & Meestringa, T. (2020). *Handboek Taalgericht Vakonderwijs. [Handbook Language Oriented Content Teaching]*. Uitgeverij Coutinho.

Harris, K. R., & Graham, S. (1996). *Making the writing process work: Strategies for composition and self-regulation*. Cambridge, MA: Brookline.

Harris, K. R., Graham, S., Brindle, M. & Sandmel, K. (2009). Metacognition and children's writing. In D. Hacker, J. Dunlosky, & A. Graesser (Eds), *Handbook of metacognition in education* (pp 131-153). New York, NY: Routledge.

Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 1-27). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Hayes, J. R. (2012). Modeling and remodeling writing. *Written Communication, 29*(3), 369-388. https://doi.org/10.1177/0741088312451260

Hayes, J. R., & Flower, L. S. (1980). Identifying the organization of writing processes. In L. W. Gregg & E. R. Steinberg (Eds.), *Cognitive processes in writing* (pp. 3-30). Hillsdale, NJ: Lawrence Erlbaum Associates.

Hillocks, G. (1986). *Research on written composition: New directions for teaching*. Urbana, Ill.: ERIC Clearinghouse on Reading and Communication Skills, National Institute of Education.

Kellogg, R.T. (2008). Training Writing Skills: A cognitive developmental perspective. *Journal of Writing Research, 1*(1), 1-26. https://doi.org/10.17239/jowr-2008.01.01.1

Klein, P. D., & Kirkpatrick, L. C. (2010). A Framework for Content Area Writing: Mediators and Moderators. *Journal of Writing Research, 2*(1), 1-46. https://doi.org/10.17239/jowr-2010.02.01.1

Koster, M., Bouwer, R., & van den Bergh, H. (2017). Professional development of teachers in the implementation of a strategy-focused writing intervention program for elementary students. *Contemporary Educational Psychology,* 49, 1–20. https://doi.org/10.1016/j.cedpsych.2016.10.002

Koster, M., Tribushinina, E. , de Jong, P. F., & van den Bergh, H. (2015). Teaching children to write: A meta-analysis of writing intervention research. *Journal of Writing Research,* 7(2). 300-324. https://doi.org/10.17239/jowr-2015.07.02.2

Langer, J. (1992). Speaking of knowing: Conceptions of understanding in academic disciplines. In A. Herrington, & C. Moran (Eds.), *Writing, teaching, and learning in the disciplines* (pp. 69-85). New York: The Modern Language Association of America.

Leijten, M., & Van Waes, L. (2013). Keystroke logging in writing research: Using Inputlog to analyze and visualize writing processes. *Written Communication 30*(3), 358–392. https://doi.org/10.1177/0741088313491692

McCarthy Young, K., & Leinhardt, G. (1998). Writing from primary document. A way of knowing in history. *Written communication 15*(1), 25-68. https://doi.org/10.1177/0741088398015001002

McCutchen, D. (1986). Domain knowledge and linguistic knowledge in the development of writing ability. *Journal of Memory and Language, 25*(4), 431-444. https://doi.org/10.1016/0749-596X(86)90036-7

McCutchen, D. (2011). From novice to expert: Implications of language skills and writing relevant knowledge for memory during the development of writing skill. *Journal of Writing Research, 3*(1), 51-68. https://doi.org/10.17239/jowr-2011.03.01.3

Moje, E.B. (2008). Foregrounding the disciplines in secondary literacy teaching and learning: A call for change. *Journal of Adolescent and Adult Literacy, 52*(2), 96-107. https://doi.org/10.1598/JAAL.52.2.1

Monte-Sano, C. (2010). Disciplinary literacy in history: An exploration of the historical nature of adolescents' writing. *Journal of the Learning Sciences, 19*, 539-568. https://doi.org/10.1080/10508406.2010.481014

Monte-Sano, C. & Allen, A. (2019). Historical argument writing: The role of interpretive work, argument type, and classroom instruction. *Reading & Writing, 32*, 1383-1410.
https://doi.org/10.1007/s11145-018-9891-0

Mottart, A., van Brabant, P., & van de Ven, P-H. (2009). Schrijven bij diverse schoolvakken. Een verkenning. [Writing in the disciplines; An exploration] *Levende Talen Tijdschrift 10*(4), 14-20. https://lt-tijdschriften.nl/ojs/index.php/ltt/article/view/181

Nokes, J. D. & De La Paz, S. (2018). Writing and argumentation in history education. In: S. Metzger & L. Harris (Eds). *International Handbook of History Teaching and Learning*. pp. 551-578. New York: Wiley-Blackwell. http://45.114.134.178:9000/digi/MP03/B1L3-0034.pdf#page=551

Rijlaarsdam, G., Braaksma, M., Couzijn, M., Janssen, T., Kieft, M., Broekkamp, H., & van den Bergh, H. (2005). Psychology and the teaching of writing in 8000 and some words. In P. Tomlinson, J. Dockrell & P. Winne (Eds.), *Pedagogy – Teaching for Learning. British Journal of Educational Psychology Monograph Series II: Psychological Aspects of Education – Current Trends*, Number 3 (pp. 127-153). Leicester: The British Psychological Society.

Schoonen, R., & de Glopper, K. (1996). Writing performance and knowledge about writing. In G. Rijlaarsdam, H. van den Bergh & M. Couzijn (Eds.), *Theories, models & methodology in writing research* (pp. 87-107). Amsterdam: Amsterdam University Press.

Schoonen, R., van Gelderen, A., de Glopper, K., Hulstijn, J., Simis, A., Snellings, P., & Stevenson, M. (2003). First language and second language writing: The role of linguistic knowledge, speed of processing, and metacognitive knowledge. *Language learning, 53*(1), 165-202. https://doi.org/10.1111/1467-9922.00213

Shanahan, T. & Shanahan, C. (2008). Teaching disciplinary literacy to adolescents: rethinking content-area literacy. *Harvard Educational Review, 78*, 40-59.
https://doi.org/10.17763/haer.78.1.v62444321p602101

Sendur, K. A., van Boxtel, C., & van Drie, J. (2021). Undergraduate L2 students' performance when evaluating historical sources for reliability. *English for Specific Purposes, 61*, 17-31. https://doi.org/https://doi.org/10.1016/j.esp.2020.08.004

Stoel, G. L., van Drie, J. P., & van Boxtel, C. A. M. (2017). The effects of explicit teaching of strategies, second-order concepts, and epistemological underpinnings on students'

ability to reason causally in History. *Journal of Educational Psychology*, *109*(3), 321-337. https://doi.org/10.1037/edu0000143

van Boxtel, C. & van Drie, J. (2018). Historical Reasoning: Conceptualizations and Educational Applications. In: S. Metzger & L. Harris (Eds). *International Handbook of History Teaching and Learning*. pp. 149-176. New York, Wiley-Blackwell. http://45.114.134.178:9000/digi/MP03/B1L3-0034.pdf#page=162

Vandermeulen, N. (2020). *Synthesis writing in upper-secondary education: From a baseline of texts and processes to process-oriented feedback*. Unpublished Dissertation, University of Antwerp.

Vandermeulen, N., Leijten, M., & Van Waes, L. (2020). Reporting Writing Process Feedback in the Classroom: Using Keystroke Logging Data to Reflect on Writing Processes. *Journal of Writing Research, 12*(1), 109-140. https://doi.org/10.17239/jowr-2020.12.01.05

van Drie, J., Braaksma, M, & Van Boxtel, C. (2015). Writing in History: Effects of writing instruction on historical reasoning and text quality, *Journal of Writing Research, 7*(1), 123-156. https://doi.org/10.17239/jowr-2015.07.01.06

van Drie, J., Groenendijk, T., Braaksma, M. & Janssen, T. (2016). *Genrespecifiek schrijven in de mens- en maatschappijvakken, Negen lesontwerpen onderzocht [Genre specific writing in the social sience subjects]*. Landelijk Expertisecentrum Mens- en Maatschappijvakken. https://www.nro.nl/sites/nro/files/migrate/genrespecifiek_schrijven_vanDrie-ea_LEMM.pdf

van Drie, J., Janssen, T., & Groenendijk, T. (2017). Improving writing in social studies through professional development: Effects on teachers' beliefs, classroom practice and students' writing. *European Journal for Applied Linguistics*, *5*(2), 273-308. https://doi.org/10.1515/eujal-2017-0009

van Drie, J., Janssen, T., & Groenendijk, T. (2018). Effects of writing instruction on adolescents' knowledge of writing and text quality in History. *L1-Educational Studies in Language and Literature, 18,* 1-28. https://doi.org/10.17239/L1ESLL-2018.18.03.08

van Drie, J., Van Boxtel, C., Jaspers, J., & Kanselaar, G. (2005). Effects of representational guidance on domain specific reasoning in CSCL. *Computers in Human Behavior 21*(4), 575-602. https://doi.org/10.1016/j.chb.2004.10.024

Van Waes, L., Leijten, M., Lindgren, E., & Wengelin, A. (2015). Keystroke logging in writing research: Analyzing online writing processes. In C.A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of Writing Research*, pp. 410-426. New York, NY: Guilford Press.

Van Waes, L., Leijten, M., & van Weijen, D. (2009). Keystroke logging in writing research: Observing writing processes with Inputlog. *GFL-German as a foreign language, 2*(3), 41-64.

Viera, A. J, & Garrett, J. M. (2005). Understanding Interobserver Agreement: The Kappa Statistic. *Family Medicine, 37*(5), 360 – 363.

Wengelin, A. (2006). Examining pauses in writing: Theory, methods and empirical data. In K. H. Sullivan & E. Lindgren (Eds.), *Computer keystroke logging and writing: Methods and applications* (pp. 107-130). Oxford: Elsevier

Wineburg, S. S. (1991). Historical problem solving: A study of cognitive processes used in the evaluation of documentary and pictorial evidence. *Journal of Educational Psychology, 83,* 73-87. https://doi.org/10.1037/0022-0663.83.1.73

Wright, K.L., Hodges, T. S., Enright, E., & Abbott, J. (2021). The relationship between middle and high school students' motivation to write, value of writing, writer self-beliefs, and writing outcomes. *Journal of Writing Research 12*(3), 601-623. https://doi.org/10.17239/jowr-2021.12.03.03

1. Appendix A: Overview of the lessons and design principles in the treatment condition and comparison condition in italics.

Lesson	Description	Design principle
1	-Introduction of the writing task -Content instruction on the situation of the Jews in the Netherlands during WW 2 -Instruction and assignments about argumentation in history (claim, argument, contra-argument, evidence based on sources) -Studying background information on the topic and making assignments focused knowledge of the main substantive concepts and on constructing a historical context (also homework.)	1.Strategy-instruction (instruction on argumentation) 3.Prewriting activities (developing background knowledge) 4.Peer interaction (assignments on background information)
2	-Instruction on the role of analysis of sources for providing evidence for a claim -Assignment: analysis of sources, using a worksheet and step-by-step-guide -Debriefing	2.Prewriting activities (analyzing sources, selecting and ordering arguments pro and contra) 4.Peer interaction (assignment analyses of sources)
3	-Instruction and short assignments on text-structure, illustrated by text examples -Teacher models text writing -Assignment: make your own working plan, based on a worksheet -Reflection: what have you learned during this lesson and what are you going to use in your own writing?	1.Strategy-instruction (instruction on text structure and modelling) 2.Studying text examples 3.Prewriting activities (workplan)
4	-Students work on the writing task based on workplan	
5	-Students read a classmate's text and give feedback based on a rubric. -Students revise their own texts	Peer review

Appendix B: Rubric used for pre- and post-task Text Quality assessment

1. GENRE-SPECIFIC ARGUMENTATION

Introduction			
1	2	3	4
None of the features listed here are sufficiently present. OR: There is no introduction.	There is an introduction and: states a **position** in the introduction. OR: introduces the subject/issue.	There is an introduction and: states a **position** in the introduction. AND: introduces the subject/issue.	States a **position** in the introduction, introduces **the subject/issue**, does **not yet provide any arguments** in the introduction and introduction encourages reading.

Body[2]			
1	2	3	4
Provides 1 or more arguments but does not elaborate these arguments or does not do so correctly.	Provides **1 relevant** argument and gives correct **proof/support for it.** OR: Provides several relevant arguments and gives correct proof/support for some of them.	Provides several **relevant arguments** and gives correct **proof/support for** all of them. OR: Provides 1 relevant argument and gives correct proof/support for it and provides 1 relevant counterargument and refutes it. OR: Provides several relevant arguments and gives correct proof/support for some of them and provides 1 relevant counterargument and refutes it.	Provides several **relevant arguments** and gives correct **proof/support for** all of them. **Orders** the arguments in the text in a systematic way. Provides 1 relevant counterargument and refutes it. The argumentation is, relatively, **convincing.**

---

[2] Focus on correct genre use, less on content. So do not judge an argument as irrelevant too strictly.

Conclusion			
1	2	3	4
A conclusion is provided, but the main position and arguments are not repeated. **OR**: No conclusion is provided.	A conclusion is provided and includes 2 of the following: □ **Main point** is repeated. □ **Main arguments** are restated briefly and clearly. □ The conclusion is **convincing**. □ No new arguments are provided in the conclusion.	A conclusion is provided and includes 3 of the following: □ **Main point** is repeated. □ **Main arguments** are restated briefly and clearly. □ The conclusion is **convincing**. □ No new arguments are provided in the conclusion.	A conclusion is provided in which **the main point and main arguments** are restated briefly and clearly. The conclusion is **convincing**. No new arguments are provided in the conclusion.

2. GENERAL WRITING

Audience orientation			
1	2	3	4
Overall, does not adopt an appealing **style** that shows reader understanding (too informal or too formal). Does not apply the **conventions** associated with the text type (argumentative letter) at all. (It is not a letter.)	Is generally inconsistent in using an appealing **style** that shows understanding for the reader. Does not apply the **conventions** associated with the text type (e.g. letter/review, essay, etc.). (It is not a letter.)	Overall, is able to maintain an attractive **style** (not too informal and not too formal) that shows understanding for the reader. Uses the conventions associated with the text type (e.g. letter/review, essay, etc.) correctly. (It is a letter, but formal conventions are not used correctly in all cases).	Overall, is able to maintain an appealing style that shows reader understanding and varies in sentence construction. Writes a reasonably catchy introduction and a fairly strong closing sentence in the conclusion. Uses the conventions associated with the text type (e.g. letter/review, essay, etc.) correctly. (It is a formal letter)

Language use & spelling			
1	2	3	4
Provides 1 or more arguments but does not elaborate these arguments or does not do so correctly	Provides 1 relevant argument and gives correct proof/support for it. OR: Provides several relevant arguments and gives correct proof/support for some of them.	Provides **several relevant arguments** and gives correct **proof/support** for all of them. OR: Provides 1 relevant argument and gives correct proof/support for it and provides 1 relevant counterargu-ment and refutes it. **OR**: Provides several relevant arguments and gives correct proof/support for some of them and provides 1 relevant counterargument and refutes it.	Provides several **relevant arguments** and gives correct **proof/support for** all of them. **Orders** the arguments in the text in a systematic way. Provides 1 **relevant counterargument** and refutes it. The argumentation is, relatively, **convincing**.

3. DOMAIN-SPECIFIC ASPECTS

Use of domain-specific concepts			
1	2	3	4
Uses no or hardly any domain-specific concepts. OR: **Makes a lot of mistakes** when using domain-specific terms.	Uses **simple domain-specific concepts** flawlessly and sometimes uses **complex domain-specific concepts**, but still makes mistakes in doing so.	Uses simple and complex domain-specific concepts correctly	Uses simple and complex domain-specific concepts and structure concepts correctly.
Content			

1	2	3	4
There is hardly any subject matter in the text. OR: Basic **subject matter** appears in the text but is largely incorrect.	**Basic subject matter** is included correctly in the text.	**Basic and complex subject matter** are included correctly in the text.	**Basic and complex subject matter** are included correctly in the text. AND: **Domain-specific relations** are made correctly or things are correctly **contextualized.** OR: There are signs of complexity and nuance (and own critical view and/or multiple perspectives).
Use of sources			
1	2	3	4
Is not always able to **summarize** information from multiple sources.	Is able to **summarize** information from multiple sources.	Is able to **synthesize** information from multiple sources (= summarize and connect) and cites **the sources.**	Is able to **synthesize** (= summarize and connect) information from various sources, cites sources and can assess the **reliability of the sources.**