

Writing in History: Effects of writing instruction on historical reasoning and text quality

Jannet van Drie, Martine Braaksma & Carla van Boxtel

University of Amsterdam | The Netherlands

Abstract: This study aims at gaining more insight in effective writing instruction to promote historical reasoning. In an experimental study, two types of instructions were compared; a general writing instruction and a discipline-based writing instruction. In addition, the effects of these instructions for students with a different initial writing ability were explored. Participants were 42 students (11th-grade), who followed a unit on the development of Dutch democracy and had to write an argumentative letter in which they argued the historical significance of a self-selected event or person. Students received a short writing instruction, based on the principle of learning from text models, in two versions: a general writing instruction or a discipline-based writing instruction. Analyses focused on historical reasoning and global text quality. Results showed a positive effect of discipline-based instruction on the quality of historical reasoning, but no effects were found on text quality. No differences were observed for good and weak writers. A pre- and post knowledge test showed improvement of students' knowledge, but no differences between conditions were found. The outcomes add to earlier studies that found positive effects of discipline-based writing instruction and provide teachers with directions for designing discipline-based writing instructions based on learning from text-models.

Keywords: writing instruction, historical reasoning, domain-specific instruction, writing-to-learn, argumentative writing



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. <http://dx.doi.org/10.17239/jowr-2015.07.01.06>

Contact: Jannet van Drie, University of Amsterdam / Research Institute of Child Development and Education, Nieuwe Achtergracht 127, Amsterdam, 1018 WS | The Netherlands - J.P.vanDrie@uva.nl

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

1. Introduction

History is a literate discipline and writing is an important means of clarifying ideas about the past, not only for historians but also in the classroom. Especially, since nowadays the focus of history education has shifted more and more towards developing students' ability in historical thinking and reasoning, instead of mainly focusing on acquiring factual knowledge of important persons, events, and dates. Students should not only learn what happened in the past, but should also be engaged in the 'how' and 'why' of history (VanSledright, 1997). Historical reasoning can be described as constructing or evaluating a description of processes of change and continuity, an explanation of a historical phenomenon, or a comparison of historical phenomena or periods. Important components of historical reasoning are asking historical questions, contextualizing, using substantive historical concepts, using meta-concepts of history (e.g., change, cause), putting forward claims supported with arguments, and using sources that give information about the past as historical evidence (Van Boxtel & Van Drie, 2013; Van Drie & Van Boxtel, 2008). Based on the writing-to-learn approach, writing can be considered an important means to engage students in historical reasoning and learning (e.g., Van Drie, Van Boxtel, & Braaksma, 2014; Voss & Wiley, 1997). However, writing in history is quite a complex activity that puts high demands on students (cf. De Oliveira, 2011; Schleppegrell, 2004). This makes it necessary to gain more insight in how to support students when writing in history.

This question came to the front in an earlier study, in which we designed and evaluated a unit on historical significance (pre-university level) that aimed to support students' historical reasoning in talk and writing (Van Drie, Van Boxtel, & Stam, 2013). The design was based on three principles to stimulate historical reasoning in the classroom: (1) open-ended tasks or questions that are meaningful from both a curriculum and a pupil perspective; (2) engaging students in historical reasoning through small group and whole-class discussions; and (3) using external representations as tools for reasoning (Van Boxtel & Van Drie, 2013). The outcomes showed that the design did enhance students' historical reasoning, but also that their writing showed weaker aspects, for example related to weighing counter arguments, contextualizing, and the use of meta-concepts (Van Drie et al., 2013). Thus, although students were able to reason historically in conversations, it seemed difficult for them to show this in writing (cf. Felton & Herko, 2004; Stoel, Van Drie, & Van Boxtel, 2015). To improve students' writing, we decided to add a writing instruction to the original unit.

Although much is known about effective writing instruction (Graham, McKeown, Kiuvara & Harris, 2012; Graham & Perin, 2007; Rijlaarsdam et al., 2011; Rogers & Graham, 2008) history teachers hardly give writing instruction to their students (cf. De Oliveira, 2011; McCarthy Young & Leinhardt, 1998), or direct this instruction only to general aspects such as explaining the requirements of a given assignment (De Oliveira, 2011). Their main concern is on covering content (cf. De La Paz & Felton, 2010; De

Oliveira, 2011). An often heard argument (at least in our country the Netherlands) is that writing instruction, as it is part of the L1-curriculum, is a task of the L1-teacher, despite the fact that research has shown that hardly any transfer occurs from what students learn about writing in L1 to writing tasks in other subjects (Mottart, Van Brabant, & Van de Ven, 2009; Zhu, 2004). In addition, the purposes of writing in L1 and other school-subjects differ. Whereas the purpose of writing in L1 is learning to write different text-genres with a focus on communicative goals, the main focus of writing in other subjects is on writing to learn subject-matter content and ways of disciplinary reasoning. Furthermore, although writing in history shares elements with writing in other disciplines, there are differences because what counts as a valid argumentation differs across disciplines (De La Paz, Ferretti, Wissinger, Yee, & MacArthur, 2012), due to the fact that the epistemological criteria for judging claims are discipline-specific (Stevens, Wineburg, Herrenkohl, & Bell, 2005). When writing in history, content knowledge and historical reasoning ability must be combined with knowledge of appropriate ways to present ideas in text. This requires the use of rhetoric and knowledge how ideas in the discipline of history can be presented (Langer, 1992; McCarthy Young & Leinhardt, 1998). This raises the question whether general writing instruction can be sufficient for promoting writing in history, or whether discipline-based writing instruction is needed. Recently, some studies found positive effects of discipline-specific reading and writing instruction (e.g., De La Paz, 2005; De La Paz & Felton, 2010; Nokes, Dole, & Hacker, 2007; Reisman, 2012). In this study we therefore compare the effects of a discipline-based writing instruction to a general writing instruction on students' historical reasoning and global text quality. Furthermore, we explore if the two writing instructions affect students with low and high initial writing ability differently.

2. Writing in history

An important goal of history education today is to develop students' ability of historical thinking and reasoning. In schools writing is often used for assessment purposes; to monitor and evaluate students' knowledge. However, several scholars in the field of learning and teaching of history consider writing also to be a means of engaging students in historical reasoning (e.g., Counsell, 1997; Greene, 1994; McCarthy Young & Leinhardt, 1998). This perspective is in line with writing-to-learn approaches (cf. Ackerman, 1993; Bangert-Drowns, Hurley, & Wilkinson, 2004; Klein, 1999) and conceptualizations of writing as a problem-solving activity (Bereiter & Scardemalia, 1987; Hayes & Flower, 1980). Scholars in the field of writing argue that during writing, students are actively engaged in the subject and explore relations among ideas, and that writing may thus contribute to constructing new knowledge, developing deep understanding, fostering conceptual change, and developing thinking skills (Klein, 1999; Tynjälä, Mason, & Lonka, 2001). Although research on writing-to-learn is often

conducted in the field of science, some positive effects have been found for the domain of history (e.g., Boscolo & Mason, 2001; Voss & Wiley, 1997).

Writing in history has discipline-specific features. When characterizing writing in history, several authors stress the role of making supported claims that are based on careful analysis of historical sources or documents (e.g., Rouet, Britt, Mason, & Perfetti, 1996; Monte-Sano, 2010; De La Paz, 2005). For example, Monte-Sano (2010) based her characterization of writing in history on the work of expert historians and described it as making a case for a particular interpretation that needs to be based on evidence and includes considering the perspective and context of the author. Although there are other writing genres in the field of (school) history, as is shown by the work of Coffin (2006), argumentative and document-based essay writing seems most dominant in schools and in research (De La Paz et al., 2012). Research suggests that this genre is, compared to writing other genres more powerful to enhance learning and understanding in history, as it engages students in knowledge transforming activities (Voss & Wiley, 1997; Wiley & Voss, 1999) and requires the use of several components of historical reasoning, for example the use and evaluation of sources, contextualization, and argumentation (Van Drie, Van Boxtel, & Van der Linden, 2006).

Although document-based, argumentative text writing in history maybe a powerful task, it is also a complex task, since it not only requires building a representation of the topic based on different single representations, but also organizing and structuring this into an argumentative structure (Rouet et al., 1996). Building an argumentative structure implies generating and ordering arguments based on one's position, and in such a way that it will convince the reader. It requires both knowledge of the content of the domain and knowledge of rhetorical processes (McCarthy Young & Leinhardt, 1998). Furthermore, the process of historical reasoning involved is complex. Research has revealed that students face various problems; use of sourcing heuristics (e.g., Britt & Aglinskias, 2002; Leinhardt, 2000; Wineburg, 1991), sound and elaborate argumentation (e.g., Leinhardt, 2000; Spoehr & Spoehr, 1994), contextualization (e.g., Van Drie, 2013), use of substantive and meta-concepts (e.g., Van Drie et al., 2013; Van Drie et al., 2006). According to De Oliveira (2011) history teachers consider the lack of development and organization of ideas a major problem. Based on interviews she found that teachers claim that students know that they are to provide examples to support their positions, but they do not know how to link the examples to the points developed in their essay or to connect them to their thesis statements. She argues that this elaboration is the most distinguishing feature between essays considered "strong" and "weak". De La Paz et al. (2012) found that better writers used more evidence, linked the evidence to the claim, contextualized and corroborated evidence and used more elaborated structures.

3. Writing instruction in history

From intervention studies (e.g., Fidalgo, Torrance, Robledo, & Rijlaarsdam, 2011; Glaser & Brunstein, 2007) and meta-analyses in writing (Graham et al., 2012; Graham & Perin, 2007; Hillocks, 1986) it becomes clear that effective writing instruction asks students to establish, and discuss clear goals and criteria for what makes a 'good text' for a certain communicative purpose and for a certain audience. Active exploration by inquiry activities of textual models that demonstrate the implementation of these criteria can be an effective learning activity (Rijlaarsdam et al., 2008, 2011). Students then construct flexible genre knowledge that could guide their content search, organization and text processing. Furthermore, students must acquire sub skills (e.g., planning, reading sources, formulating, revising) within the writing process and must learn to regulate the process as a whole. That is, students must acquire strategies: flexible sequences of actions. Active watching, comparing and evaluating teachers or peer writers at work while thinking aloud (via video's for instance) may establish schemes (metacognitive strategies) that can guide and support the monitoring of the writing process, and the reflection during and after writing (cf. Braaksma, Rijlaarsdam, Van den Bergh, & Van Hout-Wolters, 2004). The process of learning to write might be best guided in a classroom where the environmental mode (Hillocks, 1986) is implemented. In the environmental mode, the teacher sets or negotiates the task, and then organizes students' practicing of specific sub skills (such as content generation, organization, or reviewing) that contribute to the composition of the full text, with students actively discussing texts and exchanging feedback.

However, although quite a lot is known about effective writing instruction in general, only a few studies have thus far focused on effects of writing instruction in history. Monte-Sano (2008) argues that the act of writing alone is not sufficient for progression in writing in history. She compared the teaching practices of two high school teachers and their students, as one class improved on writing in history and the other did not. Effective teaching practices were among others: approaching history as evidence-based interpretation, asking students to develop and support interpretations based on evidence, using direct instruction, guided practice, independent practice and feedback. These practices are in line with earlier findings of Leinhardt (2000) and McCarthy Young & Leinhardt (1998), who conducted case-studies on the practices of teachers in relation to the progression of students on document-based writing in history. In another study of Monte-Sano (2011), the discipline-specific literacy instruction of one history teacher and the simultaneous growth in his students' historical reasoning and writing were examined. Three teaching strategies seemed to be effective: (1) annotating primary source readings; (2) regular informal writing prompts that call for a synthesis of major issues; and (3) feedback focusing on evidence use and accuracy of interpretations. She concludes that discipline-specific ways of reading and writing can help students understand history and learn to think historically, while simultaneously developing literacy skills.

De La Paz (2005) conducted an experimental study on the effects of writing instruction and historical reasoning instruction with 8th grade students with mixed writing abilities. In the experimental condition students received an instruction on historical reasoning and on writing. Students in the control group did not receive either of these two instructions. Results indicated that the students who received instruction scored significantly higher on essay length, persuasive quality, number of arguments, and historical accuracy. Building on this study, De La Paz and Felton (2010) conducted a quasi-experimental study to determine the effectiveness of an integrated reading and writing intervention on writing evidence-based arguments by 11th-grade students. The experimental group received a combined instruction on historical reasoning and written argumentation, based on a cognitive apprenticeship model. The control group received exposure to the same materials and practice in writing historical essays, without the instruction. The intervention was distributed over an entire semester. They found positive effects of the combined instruction for essay length, overall quality (overall persuasiveness and historical accuracy), number of claims, number of rebuttals, and use of documents.

These studies thus provide further support for the effectiveness of discipline-based writing instruction. They show that the cognitive apprenticeship model is a powerful model for teaching writing in history, in an intervention that is conducted over several weeks (see also a recent study of De La Paz et al., 2014). However, the question rises whether other and shorter instructional models could also be effective for teaching disciplinary writing. In this study, we will use a brief writing instruction of one lesson only, as this might be easier for history teachers to implement in their curriculum. The instruction will be based on the principle of learning from text-models (Graham & Perin, 2007; Hillocks, 1986); students are instructed to study exemplary pieces of text, along with less adequate samples, with the intention that they emulate the perceived qualities and patterns and that they develop and internalize criteria for effective writing to be used in their own writing. Two different versions of the writing instruction will be compared, a general writing instruction (as can be given in L1-writing classes) and a discipline-specific writing instruction, in which the general principles are adapted to the history domain. In addition, we want to explore the effects of the different instructions for students with different initial writing ability. In her studies, De La Paz found positive effects of the disciplinary-writing interventions for students also with lower literacy abilities (De La Paz, 2005; De La Paz & Felton, 2010; De La Paz et al., 2014). Studies of Braaksma, Rijlaarsdam, and Van den Bergh (2002, submitted) and Ferretti, MacArthur, and Dowdy (2000) show that weak and strong writers profit from different instructions. Therefore we want to find out whether students with different writing ability profit differently from the two writing instructions. Furthermore, thus far analyses of writing in history have focused solely on aspects of historical reasoning related to the use of sources and argumentation. From a domain-specific perspective it is interesting to know more about the effects on other aspects of historical reasoning as

well; for example on the use of meta-concepts and contextualization (cf. Van Boxtel & Van Drie, 2008). These aspects will be included in our analysis.

4. Aims and research questions

The aim of this study is to gain more insight in the (differential) effects of discipline-based writing instruction compared to general writing instruction on text writing in history. We also aim to add on broadening the range of instructional interventions that can be used for teaching writing in history as we incorporated learning from textual models in our intervention.

To compare the effects of a general writing instruction (WI) and a discipline-based writing instruction (DI) on text quality and historical reasoning, we used an experimental design with pre- and posttests. These writing instructions were added to the original unit on determining the historical significance of persons and events related to the development of Dutch democracy, described earlier (Van Drie et al., 2013). In addition, we explored the effects of the different instructions for students with a different initial writing ability.

The research question that guided our research was: what are the effects of a discipline-based writing instruction compared to a general writing instruction on students' (a) quality of historical reasoning in writing; (b) global text quality; (c) knowledge of the topic; and (d) knowledge of criteria for historical reasoning and argumentative writing? Our hypotheses were the following:

- students in the DI condition score higher on historical reasoning in writing;
- students in the WI condition score higher on global text quality.

Furthermore, we expect that weaker writers would profit more from the general writing instruction and score higher on global text quality and that stronger writers, who already have mastered a certain level of general writing skills, would benefit more from the discipline-based instruction and would score higher on global text quality and historical reasoning. With respect to the pre- post knowledge test, we expect no differences between the conditions on historical topic knowledge, as content-knowledge was built during the first five lessons that were the same for both groups. We assume that students in the DI condition would score higher on the knowledge of criteria for historical reasoning in writing as their intervention focused on this aspect, and that students in the WI condition would score higher on knowledge of criteria for global text quality.

5. Method

5.1 Participants

Participants of this study were 55 11th-grade students (pre-university education, 16/17 years of age); three classes and two teachers from one school. The study was conducted

over two subsequent school years; in the first year one teacher and her two classes participated, and in the second year a different teacher and her class. The students all chose history as a subject (as it is not compulsory in 11th grade) and they had not much experience with writing in history, as it is not included in the national history examinations in the Netherlands. Students that missed the pre-test writing task, the intervention lesson, or the writing task at the end of the lesson unit were excluded from the sample on which the statistical analyses were performed. The final sample contained 42 participants (22 students in WI and 20 in DI).

Students in each cohort were divided into weak and strong writers, based on a pre-test argumentative writing task (see section Instruments and analyses) on a general subject (compulsory automatic donor registration, see Appendix A) from the testing set developed by Van Weijen (2009) and Tillema, Van den Bergh, Rijlaarsdam, & Sanders (2013). Based on these scores students were divided into two groups: weak and strong writers. For this dichotomy we used the visual binning procedure in SPSS and chose for equal percentiles based on scanned cases with one cut point at 50%. Students from each group were randomly assigned to one of the two conditions before the intervention lesson took place. This procedure was carried out for each cohort. There were no significant differences in initial writing ability between the two cohorts ($F(1,40) = 1.70$; $p = .20$), neither between the two conditions ($F(1,40) = .40$; $p = .53$).

5.2 Materials

History unit. As described above, this study builds upon a previous study in which we evaluated a series of lessons that aimed at enhancing students' historical reasoning and included a writing task at the end of the unit (Van Drie et al., 2013). The lessons aimed at teaching content knowledge on political developments in the Netherlands over the last 200 years and on determining historical significance. The overarching question that guided the unit was: Which person or event was most important for the development of the Dutch democracy from 1800 till present? Determining the historical significance of people, events, and developments is a key activity of historians and includes many components of historical reasoning (Lévesque, 2008; Seixas & Morton, 2013). To enhance students' reasoning, groupwork, and whole-class discussions were included in each lesson. An overview of the lessons is presented in Appendix B. First, the students were introduced to the concept of historical significance and the criteria for establishing it, as this concept was new to the students. Afterwards, the students researched in groups one of seven historical persons or events (based on pre-selected documents). They collected, related to the criteria for historical significance, arguments pro and contra for the person/event they studied. The arguments were assembled on a sheet in a scheme. They shared their findings in short presentations with the class. Next, they formed new groups and decided on a shared ranking of different persons and events. These group-rankings resulted in a class ranking that was discussed in a whole-class discussion. In this discussion, the teacher stimulated students' thinking and reasoning by asking questions (e.g., 'Why is Universal suffrage so important?') and challenging the

students (e.g., 'What about Thorbecke, he fits all the criteria for historical significance well'). Explicit attention was given to the use of criteria for historical significance while making an argument, for example by asking to what criteria a given argument refers to.

In the writing task that was provided at the end of the unit, students were asked to write an argumentative letter to the foundation House of Democracy that planned to organize an exhibition on the development of Dutch democracy. In their letters students had to make a case for the person or event they thought was most significant for the development of Dutch democracy (see Appendix C). They could make a personal choice from the seven persons or events that were studied in the lessons. As an example we inserted in Appendix D a letter written by a student.

Writing instruction. The history unit thus focused on reasoning with historical significance. We added a brief writing instruction of one lesson of 50 minutes to the history unit described above. The information provided in the instruction was not new for the students, as it is part of their L1 writing curriculum. The aim of this lesson was to make their implicit knowledge base of writing explicit to enable transfer of their knowledge to writing in history.

We compared two different writing instructions. One that focused on general characteristics for argumentative writing (condition WI) and a discipline-based approach in which general argumentative writing was integrated with domain-specific writing characteristics (condition DI). Although the focus differed, both instructions were comparable in the types of learning activities, as they were both based on the principle of learning from text models: presenting students with text samples (Graham & Perin, 2007; Hillocks, 1986). In both conditions, the students were given an introduction by the teacher, after which they discussed in pairs several text models and afterwards were asked to formulate criteria for effective writing (see Table 1).

In Appendix E an overview of the main points of this introduction (for both conditions) can be found. In the DI-condition students received a general introduction from the teacher on argumentative text writing in history. In this instruction domain-specific aspects were discussed, such as the importance of contextualizing the topic in time, and included domain-specific examples. These aspects were not new to students as they are part of the regular history lessons, and thus also known by the students in the WI-condition. However, in the DI-condition these aspects of historical reasoning were discussed in the context of this particular assignment. Next, students were asked to study the text models in pairs. These model texts (three introductions, three main parts and three endings) were parts of original students' texts that were derived from an earlier study (Van Drie et al., 2013). However, to avoid reproduction we used texts of historical persons that were not included in this history unit (in the original unit 10 persons/events were included). The students were asked to determine weak and strong points of these texts, using the information of the teacher's introduction. Next, the students were asked to formulate criteria for effective writing, which were shared with the whole class. The teacher was instructed not to add criteria to the criteria generated

by the students. In the WI-condition students worked in the same way, but here the introduction was not adapted to the domain of history and general examples (i.e., not from the domain of history) were discussed. Afterwards, the students studied the same text models as in the DI-condition and formulated criteria for effective writing. Table 1 presents the communalities and differences of the writing instructions in both conditions.

In general, the same elements that were part of the general writing instruction were also part of the discipline-specific writing instruction, but in the latter condition discipline-specific issues were also discussed. To illustrate, in the WI students were instructed to use several arguments to support their claim and in the DI they additionally focused on what counts as a good argument in history.

Table 1. Comparison of the lessons procedures on writing instruction in the two conditions

| | Condition Writing Instruction (WI) | Condition Discipline-based Instruction (DI) |
|-----------------------------|---|--|
| Introduction by the teacher | -General theory related to the genre argumentative letter -Examples concretizing the theory related to a general subject | -Domain-specific theory related to the genre argumentative letter -Examples concretizing the theory related to establishing historical significance (different topics than in the unit) |
| Discussion of texts models | -Determining weak and strong points of text models using theory of the teacher's introduction (in pairs) -Text models included three introductions, three main parts and three endings of argumentative letters related to establishing historical significance of persons that were not included in the unit | -Determining weak and strong points of text models using theory of the teacher's introduction (in pairs) -Same text models as in WI |
| Formulation of criteria | -Formulation of criteria for effective writing in pairs -Whole-class discussion of the criteria | -Formulation of criteria for effective writing in pairs -Whole-class discussion of the criteria |

From the video-recordings of the lessons a list was made of the criteria for effective writing which the students formulated in both conditions (see Appendix F). This list shows that students were able to formulate several criteria for each text part and that they used elements of the teacher's instruction. Moreover, it appeared that the conditions differed in the formulation of the criteria concerning the main part of the

text. Students in the DI-condition formulated more discipline-specific criteria (criteria 13-16 in Appendix F) than students in the WI-condition. This suggests that the condition-specific introduction of the teacher indeed caused differences in the generation of criteria for effective writing between conditions.

5.3 Procedure

Data collection was carried out in two subsequent school years. A few weeks before the start of the unit, the students carried out the pre-test on writing. In the week before the teacher started with the unit, the students made the pre-test on historical knowledge and criteria for historical reasoning and argumentative writing. The writing post-test was carried out within a week after the writing instruction.

The own history teacher of the class taught the five lessons of the history unit. This unit was the same for both conditions. Then, the students were assigned to the experimental or control condition for the intervention lesson with the writing instruction. In the first year, the history teacher was instructed to teach the intervention lessons in both conditions. This instruction took place at the same day, but at a different timeslot. In the second year, another teacher taught the history unit. To restrict the role of the teacher in the DI-condition, the lesson was taught by the history teacher that had been involved in the first year. In the WI-condition the lesson was taught by the second author, who is an experienced language teacher and was involved in the design of this lesson. The reason for having two different teachers conduct the lesson was they had to take place at the same time (for school-organizational reasons). The lessons were given at the same time in different classrooms.

5.4 Instruments and analyses

Data included students' written texts (pre-test on a general subject and post-test argumentative letters in history), and pre- and post-tests on historical knowledge, and on knowledge of criteria for historical reasoning and argumentative writing.

Pre-test general writing task. Based on a pre-test argumentative writing task, students in each cohort were divided into weak and strong writers. The writing task was on a general topic (compulsory automatic donor registration; see Appendix A) from the testing set developed by Van Weijen (2009) and Tillema et al. (2013). The students' argumentative essays were rated on global text quality with the use of essay scales as a reference, as Schoonen (2005) demonstrated that holistic ratings (collected with essay scales) have higher generalizability than analytic scores (with scoring guides). All essays were rated by two raters (the second author and a student assistant) using a benchmark essay that was of an average text quality. The raters were provided with an extensive explanation of what aspects resulted in the average score of this benchmark essay in terms of requirements that were also specified in the instructions for the students (see

Appendix A), including text examples from the benchmark essay. This procedure served to maximize inter-rater reliability. The following aspect was for instance provided in the explanation for rating the introduction of the essays: "The introduction of the essay doesn't contain the standpoint of the writer, this is just done in the middle part. Therefore it remains unclear which standpoint the writer takes." For the middle part of the essays, for instance, the following requirements and examples were provided to the raters: "In the middle part of the essay, the writer gives three main arguments to support his opinion (a. compulsory automatic donor registration can decrease the waiting list. b. costs are low and c. there are people who are willing to be a donor but who forget to register) but these arguments aren't well supported by subordinate arguments (for instance, information is missing about the amount of people who forget to register although this information is provided in the documentation)." For training aspects and to support the raters, next to the benchmark essay and its explanation, the raters were provided with three other essays with extensive explanations of what aspects resulted in the assigned score. These essays were of a different text quality and had the following scores: 20, 95, and 150.

The raters had to award a score to each essay which expressed how much better or worse it was than the benchmark essay (cf. Blok, 1985), which was given the randomly set score of 100. If an essay was awarded a score of 200, for example, this meant that the rater thought it was twice as good as the benchmark essay. If an essay received a score of 50, it meant that the rater thought it was half as good as the benchmark essay. Raters could give all scores between 0 and infinity. However, they only used scores between 30 and 145.

As stated, all essays were rated by two independent raters. The correlation between both raters turned out to be .87 (Pearson's r), which corresponds to a reliability of .93 (Spearman Brown). For the analyses the average score of both raters was used.

Post-test argumentative letters. In the post-writing task students were asked to write an argumentative letter to the Foundation of the House of Democracy on which person or event should, in their opinion, definitely be part of an exhibition on the development of Dutch democracy (see Appendix C). The students could choose one of the seven persons/events studied during the lessons. The length of the letters had to be between 500 and 750 words and they were written on the computer. The argumentative letters were analyzed on global text quality and on quality of historical reasoning.

Global text quality was assessed with the use of essay scales using the same procedure as for the rating of global text quality of the argumentative essays in the pre-test. As benchmark essay an argumentative letter was selected that was of average text quality (which was given the randomly set score of 100). Equally to the pre-test procedure an extensive explanation was provided of what aspects resulted in the average score of this benchmark essay in terms of requirements that were also specified in the instructions for the students (see Appendix C), including passages from the benchmark essay. The letters were anonymized before scoring and the raters were

unaware from which condition a letter was coming. Two raters (the first and second author, experts in history education and writing respectively) rated all letters. Inter-rater reliability between these two raters showed to be good: the correlation between both raters was .78 (Pearson's r) corresponding with a reliability of .88 (Spearmann Brown). Also here, for the analyses the average score of both raters was used.

The quality of historical reasoning was measured using the same coding scheme that was developed for the earlier study, however slightly adapted as there were some small adaptations in the writing task (Van Drie et al., 2013). This scheme consisted of seven items that each could be scored with 0, 1 or 2 points (see Appendix G). The items were based on the framework for analyzing historical reasoning (Van Drie & Van Boxtel, 2008) and adapted to the goals of this specific task. Items included among others use of substantive and meta-historical concepts, contextualization, and argumentation in relation to the criteria for determining historical significance. Two raters (the first and third author, both experts in history education) scored the complete sample and the average score was used for further analyses. Inter-rater reliability (Cohen's κ) between the two raters (over all letters) varied between .71 and .88¹. Cohen's κ between .61 to .80 can be considered substantial and above 0.81 good (Landis & Koch, 1997). We checked whether the measures on historical reasoning and text quality were related and found no significant correlation (Pearson's $r = -.14$; $p = .38$), which suggests that separate aspects were measured. For the analyses the average score on the seven items was used.

To elucidate the different perspectives (global text quality and quality of historical reasoning) in the analyses we will discuss two examples of an introduction and two examples of the main part of students' texts. Figure 1 contains two examples of the introduction of the letter. With respect to text quality, the writer in Example 1 clarifies the motive of the letter by explaining the function of the Foundation House of Democracy. A clear standpoint is also provided. However, the introduction does not attract the reader's attention and is not very attractive. With respect to historical reasoning, there is no background information provided for Thorbecke, nor is information given about the time he lived or what is greatest merit was. With respect to general text quality, Example 2 is more effective in attracting the attention of the reader than Example 1. However, the motive for writing is not made clear and no reference is made to the House of Democracy. Nevertheless, a clear standpoint is provided. With respect to the historical aspect, a strong point is that this introduction shows some contextualization as the writer refers to the different position of women a century ago. So, Example 1 would receive a higher score with respect to general text quality at this point, compared to Example 2, and Example 2 would receive a higher score with respect to historical reasoning, compared to Example 1.

Example 1 Introduction

Dear Madam,

The House of Democracy is a place where there is only room for people and/or developments that have contributed to the democracy in the Netherlands. These include historical figures, who have meant a lot for the Dutch democracy. One of these historical figures, who has been very important for the Dutch democracy and who in my opinion should absolutely not be missing, is Johan Thorbecke.

Example 2 Introduction

Dear Madam van Dam,

The way in which we women today are living is a very beautiful thing. Men and women have been given equal rights. Also, women can have a top job at a major bank. Nowadays, we are not surprised by this, although we have a lot of respect for it. But a century ago this was very different. Men could do anything, especially the richer men. And the women? No, the women could barely do anything. But one special woman brought change to that. That woman is Aletta Jacobs. I think she is the most important person in the Dutch democracy and I would like to explain why.

Figure 1. Two examples of an introduction written by students

Figure 2 contains two examples of parts of students' arguments provided in the main parts of the texts. Example 1 is stronger with respect to text quality, and less strong on historical reasoning. The paragraph starts with the connective 'in addition', a move to inform the reader that a new argument follows. Moreover, the internal structure is made coherent by words as "this led to" and "thus". The argument is strong in convincing as the last sentence makes clear that this was important for the process of democratization. However, no direct reference is made to the criteria of historical significance. Furthermore, there is no use of meta-concepts and contextualization. Example 2 also starts with a connective: "Secondly". Detailed background information is given of the time Aletta Jacobs lived in. To clarify her historical significance, the writer makes a strong argument about the meaning of the changes Aletta Jacobs brought about for the people in her time. It is explicitly stated that she had a huge impact on the lives of women in the nineteenth century, and to us nowadays, two criteria for determining historical significance. Dates and facts are given, substantive concepts are used, however not many meta-concepts are used in this part (only 'impact'). Lastly, in the last part of the paragraph a reference is made to the discussion on the influence of

particular persons on the course of history, which is an interesting question from a historical perspective.

Example 1 Argumentation

In addition, the introduction of universal suffrage led to a larger influence of citizens in the government as well. The citizen was able to have a say in what was happening in society through his vote. From that time, he determined, together with the rest of the population of the Netherlands, who was in parliament. This led to more benefits for citizens. The party that received the majority of the votes becomes most influential in parliament. The competition between the parties increased through this, and therefore they fought harder for their principles. Thus, the voice of the majority of the population speaks through the party. This is very important in the process of democratization.

Example 2 Argumentation

Secondly, Aletta fought for suffrage of women. She wanted that women were allowed to vote, but also that women could be elected. In 1848 thanks to Johan Thorbecke a constitution was drawn. In this constitution, they actually assumed that women would not even sign up, therefore nothing had been written about the gender of the voter or the elected. In 1883 Aletta was so smart to submit a request to be included on the Amsterdam electoral list. Unfortunately, the municipality refused. In 1887 the word male was added to the constitution. Aletta was angry and tried to stand up against the old ideas about women, with as many liberal women as possible. More and more women began to recognize the importance of the right for women to vote. Even some men recognized this. Therefore, in 1884 Aletta founded the "Society of Women's Suffrage". She continued with opposition and protests, in 1919 she finally achieved where she had fought for. Women's suffrage was introduced, although women were allowed to the ballot box no earlier than 1922. This event has had a huge impact on the women of that time, but it also still of importance to us. If Aletta had not fought for this, maybe someone else had done it, but this is not to say with certainty. Who knows, maybe women today would still not be allowed to vote.

Figure 2. Two examples of main parts written by students.

Historical topic knowledge. Students' historical topic knowledge was measured in a pre- and post-test, using the same questions. The questions were related to the seven important persons and events that were studied during the lessons. For each person or event students were asked (1) to provide a brief description of the person or event, (2) to mention the period the person lived in or the event took place, and (3) to describe the relation to the development of democracy. For each of these aspects a score of 0, 1 or 2 points could be awarded, so the total score over all seven persons/events for each aspect was 14 points. The maximum score of the complete topic-knowledge test was 42 points. Two raters (the first author and a student assistant) scored all pre- and post-

tests. Inter-rater reliability (Cohen's κ) was calculated over all tests and turned out to be good: .86 (description persons/events), .87 (period) and .89 (relation to democracy). For the analyses the average score between the raters was calculated for each aspect, and the total score of all seven persons and events was used.

Knowledge of criteria for historical reasoning and argumentative writing. This test aimed to measure students' knowledge of the specific genre. Students were asked to assess (from a teachers' perspective) an essay on the quality of argumentative writing and on the quality of historical reasoning by identifying strong and weak points in the essay. These strong and weak points were related to already given aspects. For the quality of argumentative writing these aspects were: (a) the opening; (b) the middle; and (c) the closing of the text. For the quality of historical reasoning the aspects were: (a) contextualising; (b) ascribing historical significance; (c) use of historical concepts; and (d) attention for causal relation and processes of change. The essay was constructed for this particular purpose and resembled the essays students had to write on historical significance, although the topic was different (Dutch Revolt in the 16th century). We counted the correct weak and strong points mentioned for the different aspects, based on a scoring form. There was no maximum score set. Two raters (the first author and a student assistant) scored all tests. Inter-rater reliability turned out to be good: .84 for knowledge of criteria for argumentative writing and .94 for knowledge of criteria for historical reasoning (Cohen's κ).

6. Results

6.1 Effects of instruction on the quality of writing in history

The mean scores on text quality in the pretest and the posttest and the quality of historical reasoning in the posttest for both conditions are presented in Table 2. Note that the text quality scores on the pretest and posttest cannot be compared directly, as different topics and writing task are concerned. A one-way ANOVA showed a positive effect of type of instruction on the quality of historical reasoning ($F(1,40) = 5.69, p = .02$)² in students' letters (Cohen's $d = .75$, so a large effect), but not on global text quality ($F(1,40) = .02, p = .88$). Students in the discipline-based instruction scored significantly higher on historical reasoning quality than students in the general writing instruction. Additional analyses (presented in Appendix H) showed significant differences between the conditions on one subcategory of historical reasoning: use of meta-concepts ($F(1,40) = 5.82, p = .02$); Cohen's $d = .74$).

Table 2. Mean scores and standard deviations for pretest and posttest text quality and posttest historical reasoning) for both conditions

| | General Writing Instruction (WI) (N=22) | | Discipline-based Instruction (DI) (N=20) | |
|-----------------------|---|-----------|---|-----------|
| | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> |
| Pretest text quality | 95.57 | 27.12 | 100.88 | 26.97 |
| Posttest | | | | |
| -Text quality | 97.28 | 26.68 | 98.75 | 34.37 |
| -Historical reasoning | .87 | .29 | 1.08 | .27 |

6.2 Interaction effects on text quality

To examine whether the instruction had different effects for weak and good writers on text quality we performed a mixed model analysis with post-test text quality as dependent variable and condition, initial writing ability, and the interaction between condition and initial writing ability as fixed factors. No main effect of condition could be observed: $F(1,42) = 2.36$; $p = .13$. Results showed only an effect of initial writing ability on post-test text quality: $F(1,42) = 4.83$; $p = .03$, but this effect did not differ between the two conditions (condition * initial writing ability: $F(1,42) = 2.49$; $p = .12$). However, it might be that we did not find these differential effects because of the fact that the observed correlation between initial writing ability and writing ability measured in the posttest was, although significant, weak (Pearson $r = .30$; $p = .05$).

6.3 Interaction effects on historical reasoning

To examine whether the instruction had different effects for weak and good writers on historical reasoning (measured in the writing task) we performed a mixed model analysis with post-test historical reasoning as dependent variable and condition, initial writing ability, and the interaction between condition and initial writing ability as fixed factors. Results show no effects of these factors (condition: $F(1,42) = 1.57$; $p = .22$; initial writing ability: $F(1,42) = .40$; $p = .53$; condition * initial writing ability: $F(1,42) = .39$; $p = .54$). Hence, no main effects of condition and initial writing ability and differential effects of writing instructions were found on historical reasoning. However, it might be possible that we did not find these differential effects because of the fact that no significant correlation was found between initial writing ability and historical reasoning measured with the writing task that was administered after the intervention lesson (Pearson $r = -.05$; $p = 0.75$).

6.4 Effects on students' knowledge

The results of the pre- and post-test are presented in Table 3. We expected that the students would learn from these lessons and would score higher on the post-test compared to the pre-test. This expectation was partly confirmed; the students significantly improved on historical topic knowledge and knowledge of criteria of argumentative writing, but not on knowledge of criteria of historical reasoning (see Table 3). A One-way ANOVA showed no differences between the conditions on the post-test scores³ (historical topic knowledge: $F(1,37) = 1.18, p = .29$; knowledge of criteria of argumentative writing $F(1,37) = .64, p = .43$; knowledge of criteria of historical reasoning: $F(1,37) = .00, p = .98$). So, our expectation that students in WI would score higher knowledge of criteria of argumentative writing and students in DI would score knowledge of criteria of historical reasoning was not confirmed.

Table 3. Mean scores and standard deviations for the knowledge pre- and post-test and results of a t-test for paired samples ($N = 39$)

| | Pre-test | | Post-test | | p |
|--|----------|------|-----------|------|------|
| | M | SD | M | SD | |
| Historical topic knowledge | 10.71 | 5.12 | 18.69 | 7.40 | .00* |
| Knowledge of criteria of argumentative writing | 3.01 | 1.51 | 4.95 | 3.76 | .00* |
| Knowledge of criteria of historical reasoning | 2.26 | 1.30 | 2.46 | 1.65 | .50 |

* $p \leq .01$

7. Conclusions and discussion

Writing in history is a demanding activity and in order to improve students' writing in history it is important to gain more insight in the types of instruction that effectively foster discipline-specific writing. In this study, we compared the effects of two different writing instructions on general text quality and historical reasoning: a general argumentative writing instruction (WI) and a discipline-based writing instruction (DI). We expected that students in the discipline-based condition would score higher on historical reasoning in the texts and students in the general writing condition would score higher on global text quality. The first hypothesis was confirmed, the second not. We found a positive effect on historical reasoning for DI. This outcome corroborates with earlier findings and provides further support to the idea that discipline-based writing instruction can be effective for fostering historical reasoning (e.g., De La Paz 2005; De La Paz & Felton, 2010, De La Paz et al., 2014). Additional analyses of the criteria for historical reasoning showed that the students in DI improved on the use of meta-concepts. The use of meta-concepts is considered an important aspect of historical reasoning as these concepts form the core of the discipline (Limón, 2002; Van Drie & Van Boxtel, 2008) and by using these concepts the historical facts are connected and structured. The use of meta-concepts was part of the DI, however it was not

explicitly mentioned in the criteria the students came up with (see Appendix F). To gain more insight in the role of meta-concepts in text-writing, additional research is needed (for example by focusing on processes during writing in history).

With respect to the second hypothesis, we found that the WI did not score higher on global text-quality compared to the DI. This might be due to the fact that the condition only differed in the instruction the students received. After the teachers instruction students in both conditions worked with the same text models derived from historical essays. More differences between the two instructions could have led to different outcomes. An interesting question for further research would be whether students would improve on their general writing ability, which could be investigated by adding as a post-test a more general writing task (as the one used to measure initial writing ability).

Furthermore, we expected that stronger writers would profit more from the discipline-based instruction on text quality and historical reasoning, whereas weaker writers would profit more from the general writing instruction on text quality. This hypothesis was not confirmed. The observation that the type of instruction did not have different effects on text quality of weaker and stronger writers is not in line with earlier findings (Braaksma et al., 2002, submitted; Ferretti et al., 2000). This might be due to the fact that only a weak but significant correlation was found between initial writing ability and writing quality measured in the essays. Another explanation for the lack of aptitude treatment interaction might be the small sample size.

Overall, the short duration of the intervention might have been of influence on our findings. Although we deliberately choose here for a brief intervention, as most domain-specific teachers might not be inclined to pay many lessons on writing-instruction, an intervention of some more lessons might have resulted in different outcomes. It would also be worthwhile to add a delayed posttest in the design of a future study. This test should then be administered one or two months after the intervention and could show whether maintenance effects occur, in part because they have internalized the instruction more fully.

With respect to the knowledge tests, we can conclude that students in both conditions learned from the whole unit as we observed a significant improvement on the posttest compared to the pretest items related to historical topic knowledge. Furthermore, we found an improvement in knowledge of criteria for good argumentative writing. However, no improvement was found on students' knowledge of criteria for historical reasoning in writing and neither did we find any differences between the conditions on the three knowledge tests. For topic knowledge and knowledge of criteria for argumentative writing this can be due to too small differences between the conditions on these aspects. We do think that the combination of measuring students' ability in writing in history in a writing task and testing students' knowledge of the genre, by asking them to assess a text-model on various criteria, is a powerful approach. In this way one also takes into account students' knowledge about genre-specific text writing.

This study has some practical implications. First, when aiming to improve disciplinary writing, writing instruction can best be integrated within this discipline, therewith highlighting the discipline-specific aspects of writing. This could have a positive effect on students' domain-specific reasoning. Secondly, a brief writing instruction of only one lesson can lead to some improvement. This can be used as an argument for history teachers who do not want to spend so much time on writing instruction. Still, since writing in history is so demanding, it is likely that stronger effects can be expected when discipline-based writing instructions return regularly in the classroom, during and over the years, and with a focus on different writing genres (e.g., causal explanations, comparisons). Also other forms of effective writing instruction could be used, as for example peer feedback and revising opportunities (cf. Graham & Perin, 2007). Thirdly, the outcomes provide teachers with directions for designing discipline-based writing instructions. Learning from text models and developing and internalizing criteria for effective writing seem to be effective strategies. This however, does require that the teacher is well informed about criteria for good writing in history and can select adequate examples for students to work with.

To conclude, this study provides some additional support for the effectiveness of discipline-based writing instruction and points to the value of an integrated writing and historical reasoning instruction for improving students' historical reasoning. More research, with a larger sample, is needed to establish stronger effects. In addition, different designs should be developed and tested. General insights from effective writing instruction should be adapted to history and systematically investigated for different groups of students (different age and level). Furthermore, this kind of research could also be broadened to other school subjects, as geography and economics, also to gain more insight into the discipline-specific characteristics of writing in these domains.

Notes

1. Cohen's κ was respectively: criteria historical significance .88; argumentation .74; counter argumentation .74; meta concepts .77; substantive concepts .86; key-concept democracy .71; and contextualization .77. See Appendix G for a description of the items.
2. Significance level: $p \leq 0.05$
3. On pretest scores no significant differences between conditions were observed as well (historical topic knowledge: $F(1,40) = .55, p = .46$; knowledge of criteria of argumentative writing $F(1,40) = .03, p = .88$; knowledge of criteria of historical reasoning: $F(1,40) = 1.23, p = .28$)

References

- Ackerman, J. M. (1993). The promise of writing to learn. *Written Communication, 10*(3), 334-370. doi: 10.1177/0741088393010003002
- 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. doi: 10.3102/00346543074001029

- Bereiter, C., & Scardemalia, M. (1987). *The psychology of written composition*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Blok, H. (1985). Estimating the reliability, validity, and invalidity of essay ratings. *Journal of Educational Measurement*, 22, 41–52. doi: 10.1111/j.1745-3984.1985.tb01048.x
- Boscolo, P., & Mason, L. (2001). Writing to learn, writing to transfer. In P. Tynjälä, L. Mason, & K. Lonka (Eds.), *Writing as a learning tool. Integrating Theory and Practice* (pp. 83-104). Dordrecht: Kluwer Academic Publishers. doi:10.1007/978-94-010-0740-5_6
- Braaksma, M. A. H., Rijlaarsdam, G., & Van den Bergh, H. (2002). Observational learning and the effects of model-observer similarity. *Journal of Educational Psychology*, 94, 405-415. doi: 10.1037/0022-0663.94.2.405
- Braaksma, M., Rijlaarsdam, G., & Van den Bergh, H. (submitted). Differential effects of learning conditions on content knowledge acquisition and writing skills. Two experimental studies on the effects of hypertext writing and observational learning.
- Braaksma, M., Rijlaarsdam, G., Van den Bergh, H., & Van Hout-Wolters, B. H. A. M. (2004). Observational learning and its effects on the orchestration of the writing process. *Cognition & Instruction*, 22, 1-36. doi: 10.1207/s1532690Xci2201_1
- Britt, M. A., & Aglinskas, C. (2002). Improving students' ability to identify and use source information. *Cognition and Instruction*, 20(4), 485-522. doi: 10.1207/S1532690XCI2004_2
- Coffin, C. (2006). *Historical discourse. The language of time, cause and evaluation*. New York: Continuum.
- Counsell, C. (1997). *Analytical and discursive writing at key stage 3*. Longmead, UK: Blackmore Press and The Historical Association.
- 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. doi: 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. doi: 10.1016/j.cedpsych.2010.03.001
- De La Paz, S., Felton, M., Monte-Sano, C., Croniger, R., Jackson, C., Shim Deogracias, J., & Polk Hoffman, B. (2014). Developing historical reading and writing with adolescent readers: Effects on student learning. *Theory & Research in Social Education*, 42, 228-274. doi: 10.1080/00933104.2014.908754
- De La Paz, S., Ferretti, R., Wissinger, D., Yee, L., & MacArthur, C. (2012). Adolescents' disciplinary use of evidence, argumentative strategies, and organizational structure in writing about historical controversies. *Written Communication*, 29, 412-454. doi: 10.1177/0741088312461591
- De Oliveira, L. C. (2011). *Knowing and writing school history. The language of students' expository writing and teachers' expectations*. Charlotte, NC: Information Age Publishing.
- Felton, M. K., & Herko, S. (2004). From dialogue to two-sided argument: Scaffolding adolescents' persuasive writing. *Journal of Adolescent & Adult Literacy*, 47, 672-683.
- Ferretti, R. P., MacArthur, C. A., & Dowdy, N. S. (2000). The effects of an elaborated goal on the persuasive writing of students with learning disabilities and their normally achieving peers. *Journal of Educational Psychology*, 92, 694-702. doi: 10.1037/0022-0663.92.4.694
- Fidalgo, R., Torrance, M., Rijlaarsdam, G., Robledo, P., Van den Bergh, H. (2011, September). *Analysis of Instructional Components in the Strategy Instruction in Writing*. Paper EARLI conference 2011, EARLI Conference August 30- September 3, Exeter, UK.
- Glaser, C., & Brunstein, J. C. (2007). Improving fourth-grade students' composition skills: Effects of strategy instruction and self-regulation procedures. *Journal of Educational Psychology*, 99(2), 297-310. doi: 10.1037/0022-0663.99.2.297.
- 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. doi: 10.1037/a0029185

- Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology, 99*(3), 445-476. doi: 10.1037/0022-0663.99.3.445
- Greene, S. (1994). The problems of learning to think like a historian: Writing history in the culture of the classroom. *Educational Psychologist, 29*(2), 89-96. doi: 10.1207/s15326985ep2902_4
- 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, IL: ERIC Clearinghouse on Reading and Communicative Skills.
- Klein, P. D. (1999). Reopening inquiry into cognitive processes in writing-to-learn. *Educational Psychology Review, 11*(3), 203-270. doi: 10.1023/A:1021913217147
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159-174. doi:10.2307/2529310
- 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.
- Leinhardt, G. (2000). Lessons on teaching and learning in history from Pauls' pen. In P. N. Stearns, P. Seixas, & S. Wineburg (Eds.), *Knowing, teaching, and learning history. National and international perspectives* (pp. 223-245). New York: New York University Press.
- Lévesque, S. (2008). *Thinking historically. Educating students for the twenty-first century*. Toronto: Toronto University Press.
- Limón, M. (2002). Conceptual change in history. In M. Limón, & L. Mason (Eds.) *Reconsidering conceptual change. Issues in theory and practice* (pp. 259-289). Dordrecht, The Netherlands: Kluwer. doi: 10.1007/0-306-47637-1_14
- McCarthy Young, K., & Leinhardt, G. (1998). Writing from primary document. A way of knowing in history. *Written communication 15*(1), 25-68. doi: 10.1177/0741088398015001002
- Monte-Sano, C. (2008). Qualities of effective writing instruction in history classrooms: A cross-case comparison of two teachers' practices. *American Educational Research Journal, 45*, 1045-1079. doi: 10.3102/0002831208319733
- Monte-Sano, C. (2010). Disciplinary literacy in history: An exploration of the historical nature of adolescents' writing. *Journal of the Learning Sciences, 19* (4), 539-568. doi: 10.1080/10508406.2010.481014
- Monte-Sano, C. (2011). Beyond reading comprehension and summary: learning to read and write in history by focusing on evidence, perspective, and interpretation. *Curriculum Inquiry, 41*(2), 212-249. doi: 10.1111/j.1467-873X.2011.00547.x
- Mottart, A., Van Brabant, P., & Van de Ven, P-H. (2009). Schrijven bij diverse schoolvakken. Een verkenning. [Writing in various school-subjects. An Exploration.] *Levende Talen Tijdschrift 10*(4), 14-20.
- Nokes, J. D., Dole, J. A., & Hacker, D. J. (2007). Teaching high school students to use heuristics while reading historical texts. *Journal of Educational Psychology, 99*, 492-504. doi: 10.1037/0022-0663.99.3.492
- Reisman, A. (2012). Reading like a historian: A document-based history curriculum intervention in urban high schools. *Cognition and Instruction, 30*(1), 86-112. doi: 10.1080/07370008.2011.634081
- Rijlaarsdam, G., Braaksma, M., Couzijn, M., Janssen, T., Raedts, M., Van Steendam, E., Toorenaar, J., & Van den Bergh, H. (2008). Observation of peers in learning to write: Practice and research. *Journal of Writing Research, 1*(1), 53-83.
- Rijlaarsdam, G., Van den Bergh, H., Couzijn, M., Janssen, T., Braaksma, M., Tillema, M., Van Steendam, E., Raedts, M. (2011). Writing. In K.R. Harris, S. Graham, & T. Urda (Eds.), *APA Educational Psychology Handbook: Application to Learning and Teaching (Volume 3)*. (pp. 189-228). Washington, D.C.: American Psychological Association.

- Rogers, L. A., & Graham, S. (2008). A meta-analysis of single subject design writing intervention research. *Journal of Educational Psychology, 100*(4), 879-906. doi: 10.1037/0022-0663.100.4.879
- Rouet, J. F., Britt, M. A., Mason, R. A., & Perfetti, C. A. (1996). Using multiple sources of evidence to reason about history. *Journal of Educational Psychology, 88*(3), 478-493. doi: 10.1037/0022-0663.88.3.478
- Schleppegrell, M. J. (2004). *The language of schooling: A functional linguistics perspective*. Mahwah, NJ: Erlbaum.
- Schoonen, R. (2005). Generalizability of writing scores: an application of structural equation modeling. *Language Testing, 22*(1), 1-30. doi: 10.1191/0265532205lt295oa
- Seixas, P., & Morton, T. (2013). *The big six historical thinking concepts*. Toronto, Canada: Nelson Education.
- Stevens, R., Wineburg, S., Herrenkohl, L. R., & Bell, P. (2005). Comparative understanding of school subjects: Past, present, and future. *Review of Educational Research, 75*, 125-157. doi: 10.3102/00346543075002125
- Spoehr, K. T., & Spoehr, L. W. (1994). Learning to think historically. *Educational Psychologist, 29*(2), 71-77. doi: 10.1207/s15326985ep2902_2
- Stoel, G., Van Drie, J., & Van Boxtel, C. (2015). Teaching towards historical expertise. Developing a pedagogy for fostering causal reasoning in history. *Journal of Curriculum Studies 47*(1), 49-76. doi: 10.1080/00220272.2014.968212
- Tillema, M., Van den Bergh, H., 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), 1 71-97. doi: 10.1177/0265532212442647
- Tynjälä, P., Mason, L., & Lonka, K. (2001). Writing as a learning tool: An introduction. In P. Tynjälä, L. Mason, & K. Lonka (Eds.), *Writing as a learning tool. Integrating theory and practice* (pp. 7-22). Dordrecht, the Netherlands: Kluwer Academic Publishers. doi:10.1007/978-94-010-0740-5
- Van Boxtel, C. & Van Drie, J. (2013). Historical reasoning in the classroom: What does it look like and how can we enhance it? *Teaching History, 150*, 44 – 52
- Van Drie, J. & Van Boxtel, C. (2008). Historical reasoning: towards a framework for analyzing students' reasoning about the past. *Educational Psychology Review, 20*(2), 87-110. doi: 10.1007/s10648-007-9056-1
- Van Drie, J., Van Boxtel, C., & Stam, B. (2013). "But why is this so important?" Discussing historical significance in the classroom. *International Journal of Historical Learning, Teaching and Research, 12*(1), 146-168.
- Van Drie, Van Boxtel, C., & Braaksma, M. (2014). Writing to engage students in historical reasoning. In P. Klein, P. Boscolo, L. Kirkpatrick, & C. Gelati (Eds.), *Studies in Writing: Writing as Learning Activity. Studies in Writing, Volume 17, pp. 94-119*. Leiden, the Netherlands: Koninklijke Brill NV.
- Van Drie, J., Van Boxtel, C., & Van der Linden, J. L. (2006). Historical reasoning in a computer-supported collaborative learning environment. In A. M. O'Donnell, C. E. Hmelo, & Erkens, G. (Eds.), *Collaborative learning, reasoning and technology*. Hillsdale NJ: Lawrence Erlbaum Associates.
- VanSledright, B. (1997). On the importance of historical positionality to thinking about and teaching history. *International Journal of Social Education, 19*(1), 8-21.
- Van Weijen, D. (2009). *Writing processes, text quality, and task effects: Empirical studies in first and second language writing*. Utrecht: LOT Dissertation Series.
- Voss, J. F., & Wiley, J. (1997). Developing understanding while writing essays in history. *International Journal of Educational Research, 27*(3), 255-265. doi:10.1016/S0883-0355(97)89733-9
- Wineburg, S. S. (1991). Historical problem solving: A study of the cognitive processes used in the evaluation of documentary and pictorial evidence. *Journal of Educational Psychology, 83*, 73-87. doi: 10.1037/0022-0663.83.1.73

- Wiley, J., & Voss, J. F. (1999). Constructing Arguments from multiple sources: Tasks that promote understanding and not just memory for text. *Journal of Educational Psychology, 91*(2), 1-11. doi: 10.1037/0022-0663.91.2.301
- Zhu, W. (2004). Faculty views on the importance of writing, the nature of academic writing, and teaching and responding to writing in the disciplines. *Journal of Second Language Writing, 13*, 29-48. doi: 10.1016/j.jslw.2004.04.004

Appendix A: Pre-test writing task

“Compulsory automatic donor registration”: good or bad idea?

The minister of Education, Marja van Bijsterveld is organizing a national essay contest, especially for students. You’re also taking part. You absolutely want to win. The minister has appointed a jury that contains of teachers of Dutch and students from Grade 12 (pre-academic education). The winning essay will be printed in a magazine with the ten winning essays and other information on this topic. The magazine will be handed to all students from Grade 10 to Grade 12 in the Netherlands.

The subject of the essay has already been decided and was described as follows: Some politicians have recently proposed a new system for bringing in more organ donors: “compulsory automatic donor registration”. In such a system, everybody 18 years or older automatically becomes a donor after death, unless they explicitly registered that they didn’t want to. It is expected that the new system will make more organs available for transplantation. But not everybody thinks that “compulsory automatic donor registration” is a good idea. What do you think?

Assignment

Write an essay in which you give your opinion on the question: “Everybody should automatically be a donor, unless they explicitly register against it: good or bad idea?”

The essay has to meet the following requirements, set by the Jury:

1. Length: about 300 words.
2. Convince your readers (students from Grade 10 to Grade 12) of your opinion.
3. Support your standpoint well (use minimal two arguments and eventually subordinate arguments)
4. Your essay must be structured in a good and logical way.
5. Your essay must look well-cared-for (think of language use and spelling).
6. In your essay you must use at least two extracts from the ‘References’ (see next page). You must include these extracts in your essay in a meaningful way.
7. Pay attention: your readers (students from Grade 10 to Grade 12) didn’t read the references.
8. Write a title above your text.

You have 40 minutes to complete this assignment. After 35 minutes, the teacher will warn you.

Good luck!

References

Every year hundreds of people die because they had to wait for an organ for too long. For example, in the Netherlands more than 1200 people are on the waiting list for a

kidney transplant. This amounts to a waiting time of four to five years. The introduction of a system of “compulsory automatic donor registration” would be an efficient way of reducing these waiting times.

Adapted from: www.zorgkrant.nl, March 2005.

Organ donation is a form of life saving action. If you see someone drowning, you are obliged to help. In that case, choice is not an option: If you don't choose, you remain inactive, which is punishable by law. Donating organs can help and sometimes save people who are ill. This form of life saving action should be compulsory for everybody. Choosing not to be a donor, is failing to perform a lifesaving action.

Adapted from: N. Hoebe, letter to the editor, Trouw, 12th of March 2005.

Of more than eight million adults in the Netherlands, it is unknown whether or not they want to be a donor when they die. Among these eight million there are sure to be many people who would want to be a donor, but have simply never registered as such, for instance because they forgot. In a system of “compulsory automatic donor registration” the organs of all these people would suddenly become available. That's why the introduction of “compulsory automatic donor registration” is a good idea.

Adapted from: Agnes Kant, member of the Lower House for the Socialist Party, www.zorgkrant.nl, March 2005.

The more organ donors, the better. But always out of free will! Compulsory automatic donorship, unless you explicitly declare yourself against that, is not a good thing. “Silence is consent” is not a proper way of reasoning. It is a clever way of abusing the lack of knowledge of many people. Suddenly many more organs will become available, ripped from ignorant civilians, such as the homeless, who were not aware of becoming a donor after they died. And all of this under the pretext of “silence is consent”.

Adapted from: 'Dave95' on the internet forum on www.debatplaats.vara.nl, March 2005.

“Until what age do my organs qualify for donation? How healthy do I need to be? Under what circumstances do I need to die? And most important of all: what is brain dead? It would be highly unpleasant if, during a near-death experience, I would see my organs being removed from my body! These are the questions that nobody is able to answer for us, especially not the Minister of Welfare and Health. Until then my answer will be ‘no.’”

Source: C. Putter, letter to the editor, Trouw, 12th of March 2005.

“How often do we have to say that people don't die because they didn't receive an organ, but because they were critically ill. In my view, “compulsory automatic donor registration” is undesirable.”

Adapted from: Th. van der Kraats, letter to the editor, Trouw, 12th of March 2005.

Appendix B: Description of the lessons

Introduction

In 2013 the House of Democracy will be opened in the Hague (the Netherlands), The aim of the House of Democracy is to inform the people and especially the youth about democracy and how it functions in daily practice.

In this lesson unit you study how democracy developed in the Netherlands. Together with your classmates you examine a few persons and events that could get a place in the House of Democracy. At the end of this unit you will write a letter of recommendation to the Foundation House of Democracy, in which you make a case for one person or event which you think should definitely have a place the House.

Overview of the lessons

Lesson 1: Introduction

Introduction to the concept of historical significance and making a list of criteria to decide on historical significance.

Lesson 2: Building a context

Each group studies their person or event. Collaborative construction of a time-line of the period 1800 – present, with the whole class.

Lesson 3: What is the significance of your person/event?

Each groups collects arguments from pre-selected sources and orders the arguments to the criteria of historical significance.

Lesson 4: One-minute recommendations in front of the class

Each group has one minute to make clear to the rest of the class what the significance is of the person/ event they studied.

Lesson 5: The class top 10

New groups are composed. In these groups the students compose their group top 10. This results in a kind of class election, so finally there is a class top 10. This top 10 is discussed in a whole-class discussion.

Appendix C: Post-test writing task: Argumentative letter

In the history lessons you have discussed the historical significance of seven persons and events related to the development of democracy in the Netherlands. Now you can make your own choice of who/what was most important and support that choice with arguments. You have to write an argumentative letter to the secretary of the House of Democracy. With your letter you want to convince the secretary of the historical significance of the person or event of your choice.

1. Write an argumentative letter to the secretary of the House of Democracy in which you make clear which person or events you think has been most significant for the development of Dutch democracy and therefore definitely should be part of the exhibition.
2. Choose from the list of persons and events the one you think is most significant.
3. Provide arguments why you think this person/event should receive a lot of attention in the exhibition.
4. The length of your letter is between 500 and 750 words.
5. You may make use of the information provided by the teachers (sources that were studied in the previous lessons).
6. You have 100 minutes to accomplish this task.
7. Don't forget to mention your name, the name of the school and the date.

Good luck!

Appendix D: Example of a student's argumentative letter (anonymized and translated from Dutch)

Democracy street 26
1965 XP the Hague

31-01-2012, at City

Dear commission of the House of Democracy,

Once, there was a young man named Johann. He looked upon Dutch politics with sorrow and considered it to be time for radical changes. The changes had to be liberal, so everyone would be free in their conduct. This young liberal was none other than Johann Thorbecke, who according to some people, including myself, has been the biggest political reformer of the Netherlands. Therefore, it would be only logical if he would be granted a position in the prestigious House of Democracy, worthy of his importance. Now, you will probably ask why he deserves this position and what this man has done to earn that place. I will try to explain this as much as possible.

Johann Thorbecke has implemented good reforms from which contemporary Dutch society still benefits. He has ensured that the power in Netherlands became the responsibility of the ministers and no longer of the king. This was a scoop in Europe, because after the French Revolution monarchies had been restored everywhere and the kings regained all power. Thorbecke was of the opinion that this should change, because some kings ruled by arbitrariness and the Dutch people of that time were not happy with the king. Instead of launching a violent rebellion, Thorbecke was able to convince the king of his right. As a result, the Netherlands switched peacefully into a constitutional monarchy. This was also a scoop: never before had such a major change been made without bloodshed.

Thorbecke was also very good for the development of the Dutch trade. For instance, he allowed the construction of railways and asked for the digging of the North Sea Canal and the New Waterway. This allowed the Netherlands to develop into what they are today: the most important transit port of Europe. Furthermore, this allowed Rotterdam to develop into one of the major ports of Europe. The shipping trade in the Netherlands then and in contemporary society, are therefore partly due to Johann Thorbecke.

Furthermore, Thorbecke stimulated the interest in Dutch politics. This allowed for voting in the Netherlands to become a right, instead of an obligation. If he wouldn't have done that, there would have been far less people who went to vote and Dutch cabinets would not have been elected by the majority of the people, but maybe only by 12 percent. He did this partly by introducing the right to vote in the Netherlands.

A widely used counter argument is that the right to vote was not universal, but that it was a census suffrage only opened to men. Census suffrage means that only the citizens who pay a certain percentage of tax are allowed to vote. This is absolutely true. Unfortunately, these people do not understand from which starting point the Netherlands had developed in terms of democracy.

Before this law was introduced, absolutely no one had active participation in politics therefore the Netherlands had been not democratic at all. And Thorbecke introduced the law in this way with a good reason, considering the time. In that time only the richest, those who could pay the tax rate, went to school. He held the opinion that to be able to make such an important choice meant that it was necessary to be well educated. Therefore, he allowed only the tax payers to vote because they had enjoyed education and could form a better opinion. In addition it was only logical at that time that only men were allowed to vote, because women could do nothing according to the people.

Thorbecke is the most important political reformer of the Netherlands. He conceived parliamentary democracy, flourishing seaports, and census suffrage, which was the biggest step towards democracy. These are all developments that the Netherlands still enjoy and that people back then could benefit from. Furthermore, these are laws that paved the way for the ensuing democratic steps and therefore, were of great influence.

For these reasons, there is no doubt that Thorbecke should obtain a central position in the House of Democracy because his laws have made him a symbol and he is nothing less than the greatest political reformer of the Netherlands.

Sincerely Lowyck from the XXX Lyceum, class XXX

Street, City

Appendix E: Overview of the main points of the teacher's introduction for both conditions

| Condition General Writing Instruction (WI) | Condition Discipline-based Instruction (DI) |
|---|---|
| Explanation and definition "argumentative text" with reference to the L1 writing lessons. | Idem ditto |
| Elements of introduction of an argumentative text: 1) attention getter, 2) topic and issue, 3) standpoint of the writer. | Elements of introduction of an argumentative text: 1) attention getter, 2) topic and historical context, 3) issue, 4) standpoint of the writer |
| Caveats when writing the introduction: 1) no topic / issue, 2) argumentation in introduction, 3) too much information. | Caveats when writing the introduction: 1) no topic / issue, 2) argumentation in introduction, 3) too much historical information. |
| Elements of the main part (second and following paragraphs): 1) arguments and subordinate arguments, 2) refutation of counterargument, 3) use of connectives, 4) potential extra information. | Elements of the main part (second and following paragraphs): 1) arguments based on criteria for historical significance, 2) subordinate arguments with evidence, 3) refer to sources, 4) refutation of counterargument, 5) correct use of historical concepts, 6) use concepts as change, cause etcetera and make correct historical relations. |
| Caveats when writing the main part: 1) arguments are too similar, 2) subordination is incomplete or unreliable, 3) advice: mention sources, 4) advice: think of the order of the arguments (end with strongest?!) | Caveats when writing the main part: 1) no use of criteria for historical significance, 2) historical changes and explanations are not explicitly mentioned and elaborated, 3) no use of historical context, 4) subordination is historically incorrect or unreliable, 5) advice: use concepts that belong to history. |
| Elements of ending (last paragraph): 1) end your text, say goodbye to your reader, 2) repeat your standpoint in different words, 3) summarize your arguments, 4) use a strong closing sentence. | Elements of ending (last paragraph): 1) end your text, say goodbye to your reader, 2) repeat your standpoint in different words, 3) summarize your arguments with refer to historical significance, 4) use a strong closing sentence. |
| Caveats when writing the ending: 1) ending is too long, 2) no repetition of standpoint, 3) | Idem ditto. |

| Condition General Writing Instruction (WI) | Condition Discipline-based Instruction (DI) |
|--|---|
| formulation of new arguments, 4) "text goes out as night candle" (no attractive ending) | |
| Conventions of a letter with reference to the L1 writing lessons: name and address of addressee, date, reference beginning (dear ...), ending, layout. | Idem ditto. |

Appendix F: List of criteria for effective writing formulated by students for both conditions

| | General Writing Instruction (WI) | Discipline-based Instruction (DI) |
|--------------|--|--|
| Introduction | 1. use an engrossing start, quotation or anecdote (introduction should have 'sex appeal') 2. give a clear standpoint / opinion 3. be complete, not too short, mention all components 4. correct addressing, polite salutation 5. mention the issue 6. explain the topic 7. do not use arguments | 1. use an engrossing start, quotation or anecdote 2. give your standpoint 3. be complete, not too short, mention all components 4. address, salutation 5. write down the reason for the letter 6. - 7. - |
| Main part | 8. clearly elaborate your arguments 9. use connectives 10. refer to sources and use them to make your arguments stronger 11. well-structured paragraphs / well-structured sentences 12. mention counter arguments and refute these 13. refer to criteria of historical significance 14. use historical concepts 15. it should be historically correct 16. provide a historical context 17. - 18. - | 8. clearly support your arguments 9. take care of connectives 10. refer to sources 11. well-structured paragraphs 12. mention counter arguments and refute these 13. - 14. - 15. - 16. - 17. write persuasively 18. use historical facts |
| Ending | 19. provide a short summary with the main arguments 20. repeat your standpoint 21. strong closing sentence 22. strong conclusion 23. end politely 24. pay attention to the length in relation to the main part | 19. provide a short summary with the main arguments 20. repeat your standpoint 21. strong closing sentence / engrossing ending (something personal) 22. strong conclusion 23. - 24. - 25. do not mention new arguments |

25. -

Appendix G: Assessment of quality historical reasoning of the argumentative letters

| | 0 point | 1 point | 2 points |
|--|--|--|--|
| A. Reference to different criteria for establishing historical significance (important in the time itself, important for the present times, is a symbol for) | 1 criterium provided | 2 criteria provided | 3 criteria provided |
| B. Use of elaborate argumentation to support claim | Hardly any argumentation or not correct | Argumentation is limited | Argumentation is correct, elaborate and adequate |
| C. Counter argumentation (arguments against the historical significance a of person or event chosen are given and refuted) | No counter argumentation or not correct. | Counter argumentation, but not convincing and/or not refuted | Counter argumentation and rebuttal are convincing |
| D. Explicit use of meta-concepts (e.g., cause, consequence, change, impact) | No use, or no correct use of meta-concepts | Limited use of meta-concepts | Adequate use of meta-concepts |
| E. Use of substantive concepts | No use or no correct use of substantive concepts | Limited use of substantive concepts | Adequate use of substantive concepts, including explanations |
| F. Argumentation connected to the key concept democracy | Not connected to the concept | Connected, but in a limited way | Adequate connection |
| G. Person/event is contextualized in time | No or incorrect contextualization | Limited contextualization | Adequate contextualization |

Appendix H: Descriptive statistics and test statistics for posttest sub scores historical reasoning for both conditions

| | General Writing Instruction (WI) (N=22) | | Discipline-based Instruction (DI) (N=20) | | <i>F</i> | <i>p</i> |
|--|---|-----------|---|-----------|----------|----------|
| | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | | |
| - Historical significance | 1.02 | .79 | 1.30 | .59 | 1.61 | .211 |
| - Argumentation | 1.10 | .45 | 1.28 | .55 | .36 | .550 |
| - Counter argumentation | .57 | .64 | 1.00 | .84 | 3.53 | .068 |
| - Meta-concepts | .50 | .67 | 1.00 | .67 | 5.82 | .021* |
| - Substantive concepts | 1.11 | .60 | 1.25 | .55 | .59 | .447 |
| - Connection to key concept democracy | .89 | .83 | .98 | .80 | .12 | .727 |
| - Contextualization | .80 | .75 | .78 | .64 | .01 | .925 |

* $p \leq 0.05$