



UvA-DARE (Digital Academic Repository)

Disciplinary writing

Four empirical studies on historical and philosophical literacy

Holdinga, C.C.

Publication date

2023

[Link to publication](#)

Citation for published version (APA):

Holdinga, C. C. (2023). *Disciplinary writing: Four empirical studies on historical and philosophical literacy*. [Thesis, externally prepared, Universiteit van Amsterdam].

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

CHAPTER 4

IMPLEMENTING WRITING PROCESS INSTRUCTION IN THE HISTORY CLASSROOM: A QUASI-EXPERIMENTAL STUDY*

In this study, we have investigated if integrating writing tasks as learning activities within the history classroom, when accompanied by process instruction, would foster students' historical writing proficiency, epistemic experience of writing, and course content knowledge (grades 10-12). We conducted a quasi-experimental study with two experimental conditions ($N = 182$) and a control condition ($N = 86$). Teachers participating in the experimental conditions attended a professional development session to design evaluative writing-to-learn tasks, tailored to their own curriculum. In both experimental conditions, teachers replaced textbook assignment with their new-designed evaluative writing-to-learn tasks. In one experimental condition, teachers added strategy instruction, to support students to perform tasks step-by-step, focusing on disciplinary aspects of inquiry and text. The control condition was a non-writing condition.

Significant effects in a transfer task of the additional strategy instruction were seen in all aspects of text quality, holistically and on discipline-specific criteria, and in epistemic experience. Recall test results showed a similar increase in course content knowledge for the experimental conditions, compared to a non-writing condition. These results imply that writing seems a promising alternative activity for history learning, yet this only results in more advanced writing proficiency when attention is paid to the writing process, by means of strategy instruction.

* Chapter 4 is based on: Holdinga, C. C., Van Drie, J. P., & Rijlaarsdam, G. C. W. (2023). Under-rated yet effective: Writing process instruction in the history classroom. Under review.

This study underscores the potential of reading-thinking-writing process instruction in the content classroom, as a contribution to the development of students' disciplinary literacy skills, without losses for content learning.

1. INTRODUCTION

For every student to become literate – that is one of education's social tasks. In secondary education in the Netherlands, this is an explicit formal goal (Dutch Education Inspectorate, 2022). The idea behind this goal, is that a literate person is more likely to succeed in further education, and in the labour market; moreover, literate individuals are better able to communicate their knowledge. The call for a sound literacy-base has become more urgent now that students' proficiency level seems to decline, according to research of the Dutch Education Inspectorate (2022). Based on these results, the improvement of students' basic skills (reading, writing and mathematics) has been put on the agenda of every school leader.

According to Dutch educational policy, the development of students' literacy is a shared responsibility: all teachers should contribute (Dutch Ministry of Education, Culture, and Science, 2009). However, this is not a shared opinion among teachers. Language teachers are the ones who primarily take responsibility for developing students' literacy, even though language-oriented teaching in subjects has proven its value. From previous studies, it is clear that subject teachers' effort to support students' reading and writing within the disciplines, has additional value to the development of students' literacy (see for example Yore, Bisanz, and Hand's (2003) overview of literacy studies in science). However, in most subjects, the culture of teaching is still primarily focused on subject-specific content knowledge and reasoning. The belief that it is important for a student to be able to communicate content knowledge or reasoning, seems to be underrepresented (Siebert & Draper, 2008).

Biancarosa and Snow argued that "teachers need to realize they are not just teaching content knowledge, but also ways of reading and writing specific to a subject area" (Biancarosa & Snow, 2006, p. 20). To be able to become a subject-area expert, it is necessary to acquire discipline-specific reading and writing heuristics (Monte-Sano et al., 2014; Goldman et al., 2016). And, maybe even more important, it also works the other way around: the development of disciplinary literacy strengthens thinking and reasoning within that discipline. Literacy skills "can act as vehicles for deeper content acquisition" (Miller

et al., 2018, p. 84). Teaching literacy and teaching content do not conflict, they complement and strengthen each other.

Nonetheless there are obstacles. A first obstacle, is that dominant subject area norms often get in the way (Moje, 2008; O'Brien, 1995). The "pedagogy of telling" (see O'Brien et al., 1995) is still dominant in subject teaching; teachers feel pressured to transmit subject content in a timely and effective manner. This results in the view that supporting students in processing source texts and writing texts, is only a time-consuming, additional, burden, in an overloaded instructional agenda (Mac Mahon, 2014).

Teaching literacy is not only felt as a burden, it is also seen as a responsibility for others. Moje (2008) premised that "it is not uncommon to hear teachers in [...] subject areas argue that they should not be expected to assess a student's ability to construct a well-argued essay for their class: 'What matters is the content,' they say, 'I'm not the [language] teacher'." (Moje, 2008: 98). Many teachers initially reject the idea that they are pre-eminently the teachers to highlight the conventions of literacy within their disciplines. Providing writing instruction is thus generally highly underrated by subject teachers.

A second obstacle is knowledge. Teachers are not commonly skilled to teach disciplinary literacy, they feel they have little knowledge about reading and writing processes, and little experience on how to support writing (De Oliveira, 2011). Teachers, although being proficient readers and writers, who are familiar with the academic discourse, may lack awareness of the mental processes needed to interpret texts in their disciplines (Alger, 2007). Consequently, teachers have spent little or no time demonstrating or modeling explicitly how to read and write in their particular subject area. Additionally, preservice teachers are sceptical about the efficacy of teaching disciplinary literacy, offered by disciplinary literacy research. They therefore tend to avoid implementation of such practices (Moje, 2008). However, that is if literacy instruction is even a vital part of preservice teacher training at all, which is often not even the case (De Oliveira, 2011).

In the present study, we aimed to overcome these aforementioned obstacles. What if teachers would make more use of writing as a learning activity in their lessons? And, what if they were supported to provide students with writing process support? In the design study of Chapter 3, we have developed a discipline-specific writing strategy instruction. In the current study, we added a professional development program aimed at teaching how to design writing tasks, to implement writing instruction, and to support students' discipline-specific writing process, with the overall aim of improving students'

disciplinary source-based writing. We focused on students in pre-university level, upper secondary school (grades 10-12, age 15-18), as upper secondary school levels provide opportunities to work with more extensive writing tasks concerning more complex issues, aimed at profound understanding of content. This way, reading, writing, and content learning, would be intertwined. Although the obstacles mentioned might seem surmountable issues, they are deeply rooted in teachers' and students' routines. Therefore, in the current study, we tried to adapt to the existing classroom routines in two ways. First of all, teachers developed writing tasks tailored to their own curriculum. This ensured ecological validity. We considered this a strength of the study; if the intervention turned out to be effective, it might be in other situations as well, with other teachers, in other schools. Secondly, following Moje (2008), in our study we have tried to foreground the discipline. In our case, this was the subject of history, which is known for its persistent subject culture of "content first". As Gillis (2004) advised, we did not adopt generic reading and writing strategies into the history classroom, but we adapted reading and writing strategies into a discipline-specific strategy for historical writing based on historical sources. Source-based writing is a common task in history classes in upper secondary grades, yet linguistically demanding for students.

1.1 Linguistic Challenges in History

The school-subject history comes with many linguistic challenges for students, concerning both reading and writing. First of all, textbooks contain multiple text types: informative texts which are full of dense and abstract language, and source texts, especially primary sources, which are often difficult to read, and open to interpretation. Students are expected to read and assess historical sources of different kinds: these might be texts, but can also be cartoons, photos, or graphics. Sourcing is an important skill to develop in secondary school. When multiple documents are involved, it is especially important for students to attempt to contextualize the information from each document, to connect information elements from different texts, and to create a coherent, integrated model necessary for understanding (Wiley et al., 2014). Furthermore, students must be able to discuss complex issues in oral or written form, based on source inquiry. Although several text genres common to history can be distinguished (Coffin, 2006), the most prominent is argumentative writing: "Argumentative historical writing, through which historians defend their interpretations, their use of evidence, their research methodologies, and the significance of their work, represents the pinnacle of historical

writing" (Nokes & De La Paz, 2018, p. 559). Reading and writing thus play a significant role in the learning of history, however, students experience these to be challenging (Schleppegrell & De Oliveira, 2006; Goldman et al., 2016).

Although writing is not grounded as a learning activity within the history classroom in the Netherlands, it has proven its great potential for learning in general (Graham et al., 2020), as well as for learning history specifically (Voss & Wiley, 1997; Wiley & Voss, 1996, 1999). Previous research of Monte-Sano (2008) has shown that simply adding writing into content lessons is not sufficient for growth in historical writing. Strategy-instruction on reading strategies (Nokes et al., 2007; Reisman, 2012; Van Driel et al., 2022b), and reading-writing strategies (De La Paz, 2005; De La Paz & Felton, 2010; De La Paz et al., 2014) have proven to be effective for the development of historical reasoning. Therefore, the aim of the present study was to improve students' writing proficiency through (1) inclusion of source-based writing as a learning activity, and (2) a discipline-specific writing strategy intervention in the history classroom, considering the obstacles encountered in practice.

1.2 Research Questions

Our research questions were:

RQ1: To what extent does the replacement of workbook assignments by evaluative writing tasks contribute to (a) disciplinary source-based writing proficiency, (b) epistemic experience of writing, and (c) students' course content knowledge?

RQ2: To what extent does the addition of strategy instruction to such writing tasks contribute to (a) disciplinary source-based writing proficiency, (b) epistemic experience of writing, and (c) students' course content knowledge?

We hypothesized that the replacement of workbook assignments by evaluative writing tasks, as knowledge constructing learning activities (RQ1), would not result in higher-quality texts (Monte-Sano, 2008), but that that it might result in more epistemic experience, a term derived from Bereiter (1980). With the epistemic experience of writing, we refer to the awareness of acquiring and constructing knowledge through writing. We hypothesized that students who would perform evaluative writing tasks as learning activities, would experience the epistemic function of writing. Therefore, also in transfer tasks, we expected that students would experience writing as a learning activity,

resulting in a greater epistemic experience, more than students in a non-writing condition.

Furthermore, we predicted that students would gain more content knowledge of the topics they wrote about, by means of more extensive writing tasks, than by other (non-writing) learning activities, since the act of writing has the potential of developing a more deep and complex knowledge (Graham et al., 2020).

Regarding RQ2, we expected students who received strategy instruction to write higher quality texts (Graham & Perin, 2007), than students who only performed writing tasks without strategy instruction, or students in a non-writing condition. We also expected the strategy instruction to reinforce effects on epistemic experience, since the strategy instruction would ease the writing process, and therefore create more space for content development. Hypothetically, for students who received strategy instruction, this would occur both in transfer tasks, resulting in greater epistemic experience, as well as in the writing tasks used as learning activities, resulting in more course content knowledge.

To examine whether effects of conditions interacted with learner characteristics, we added a third, explorative research question:

RQ3: To what extent are effects on RQ1 and 2 moderated by learner characteristics (writing beliefs and self-efficacy)?

We aimed our lesson materials to be equally effective for all students, regardless of belief types or self-efficacy levels with regard to writing. However, learner variables could moderate intervention effects. Previous research on writing beliefs from White and Bruning (2005) has defined two main belief types for writing: transmissional beliefs, which are associated with the perception that writing is mainly a means of transmitting information to one another; and transactional beliefs, which are associated with the conception that writing might fulfil an epistemic function. Spanish researchers concluded that students with high transactional and low transmissional writing conceptions, might be expected to write higher quality texts (Villalón et al., 2015). We thus hypothesized that students with high transactional beliefs would write higher quality texts on all measurement occasions, and would experience writing more often as an epistemic experience.

A second learner variable, which might have direct effects, and also moderate the effects, is students' self-efficacy. Previous research has shown self-

efficacy to be positively associated with writing performance (Pajares, 2003). In our study we aimed to explore the possible moderations of self-efficacy, by assessing if, for example, students with low self-efficacy for writing would benefit more from the strategy instruction, since they potentially require more hold for the writing process, or if this works inversely.

2. METHOD

2.1 Research Design

We implemented a quasi-experimental pre-test/post-test design, with three conditions: two experimental conditions: Writing Task and Strategy (WT+S), and Writing Task (WT), and a Control Condition (C). In the WT+S Condition, the teachers attended a professional development (PD) session about epistemic writing task-design. In order to replace textbook assignments, they designed such writing tasks themselves, under researcher's guidance and monitoring, and implemented a writing strategy instruction. In the WT Condition, the teachers also attended the same PD-session, but they were not informed about the strategy instruction, and did not implement it. In the Control Condition, the teachers provided regular history education, without evaluative writing tasks or strategy instruction. In all conditions, the teachers taught the history content as pre-planned in the school specific year plan, which varied between groups. We considered this a strength of our design; it enhanced ecological validity, and generalizability across thematic contents. An overview of the research design is presented in Table 4.1.

2.2 Participants

In this study, 11 qualified history teachers, from 10 different schools, participated with 14 groups. Initially, 12 teachers participated in this study, with 16 groups, however, in two groups taught by the same teacher, data collection procedures were not executed according to established criteria. Therefore, these two groups were excluded from further data collection and subsequent analysis.

Since motivation of the teacher was considered an important factor for successful educational innovation, teachers were offered a choice with regard to their preferred experimental condition, i.e, WT+S or WT. Subsequently, other teachers were engaged to participate in the Control Condition; these teachers received the experimental lesson materials following the post-test, for them to potentially implement in their subsequent lessons, if preferred.

Table 4.1. Research Design

Condition	N = 268	PD session	Pretest	Intervention		Posttest
				Writing tasks	Strategy instruction	
WT+S	n = 119 7 groups 6 teachers	+	+	+	+	+
WT	n = 63 3 groups 2 teachers	+	+	+		+
Control	n = 86 4 groups 3 teachers		+			+

Note: When an activity was conducted, a + is presented.

The students participating in this study were enrolled in the pre-university program (grades 10-12). Group-size differed from 10 to 32 students ($\mu = 20.6$). Due to Covid-19 restrictions (quarantines or isolations), the rate of students' absence was higher than anticipated, and, due to this, insufficient data was available for 10 students. All students – and in the case of minors, also their parents – actively consented for participation. When consent was withheld (by either students or parents), students did participate in the classroom activities during the intervention period, since these replaced regular activities, but their data were not collected. Overall, from ten students (WT+S: 2; WT: 4; C: 4) no consent was obtained; there were no parents who objected to participation of their child.

The results presented in this study were based on the data of 268 students (14-19 years, $\mu = 16.2$, $sd = .85$).

2.3 Procedures

2.3.1 Professional Development Session

The nine teachers who taught the experimental conditions (WT+S and WT), took part in an online one-hour PD session, about teaching writing in history, and specifically, writing task-design. In this session, two researchers explained and discussed the background of the study, the main ideas behind the lesson materials, and four, hereafter described, writing task design principles.

Principle 1: it is advised to make use of evaluative questions, which ask for inquiry of different perspectives on an issue. Such tasks have the potential to develop understanding of content (Newell & Winograd,

1995; Voss & Wiley, 1997) and to elicit historical reasoning (Monte-Sano & De La Paz, 2012; Van Drie et al., 2006).

Principle 2: writing from sources is advised, since reading and writing may reinforce each other (Graham et al., 2018; Wiley & Voss, 1996). For historical writing, it is best to preselect, preferably, primary sources, representing multiple perspectives on the issue presented in the evaluative question (Britt & Rouet, 2012).

Principle 3: the writing task must replace regular workbook assignments, and are not 'extra'. This way the writing task becomes an actual writing-to-learn task, instead of a simpler processing, or reproduction task, of smaller content units.

Principle 4: the writing task can be completed within one lesson hour. Rationale behind this principle was that students' writing processes become visible for the teacher, since the writing would take place in the classroom.

The teachers themselves designed the writing tasks based on these four principles, to use during the intervention period. They tailored the tasks to their own school-specific year plans, with regard to the content of the tasks. The research team supported the teachers by providing written or oral feedback on draft versions. Generally, it was advised that each teacher developed two writing tasks for their units. An overview of topics and evaluative questions, is presented in Table 4.2. To assess the extent to which the developed tasks complied with the design principles, all tasks were coded (Table 4.3). As presented, teacher C developed only one writing task, which was more extensive, and was without preselected sources, because they felt it suited their students' grade (12th) better to let each search for appropriate sources themselves. It was decided that all developed tasks, including the task of teacher C, were of high enough quality to be included in further analyses.

Table 4.2. Overview of Topics and Tasks

Condition	Teacher	Group	Grade	Topic during intervention period	Evaluative question (Task ID)
WT+S	1&2	1&2	11 th	Modern imperialism related to industrialization	To what extent did Industrialization lead to a higher standard of living? (A1) To what extent was the ethical politics experienced as a continuation of the Dutch domination of the Dutch East Indies? (A2)
	3	3&4	10 th	The expansion of religions in the Middle Ages	How important was the battle of Poitiers in halting the advance of the Saracens? (B1) To what extent were crusaders guided by their own interests when participating in the crusades?(B2)
	4	5	12 th	Decolonization	Open assignment: students formulated their own inquiry question and searched for suiting sources themselves. (C)
	5	6	11 th	The conflict in the Netherlands that resulted in the foundation of a Dutch state	To what extent was the relationship between Philip the Good and Ghent's citizens around 1450 illustrative for the relationship between princes and city councils in the same period? (D1) To what extent was the Iconoclasm of 1566 an expression of religious, political and/or socio-economic discontent? (D2)
	6	7	11 th	The Age of Enlightenment	To what extent is it justified to introduce compulsory vaccination in a pandemic, starting from ideas of John Stuart Mill? (E1) May a Prime Minister call upon citizens to have their children vaccinated against COVID, if you take as your starting point the ideas of the Enlightenment philosophers Locke and Rousseau? (E2)
	WT	8	9&10	11 th	The division of the world into two ideological blocs
9		11	11 th	The Soviet Union	Which of the causes of World War I do you think was most important? (G1) To what extent was the Soviet Union a workers' paradise in the period 1924-1941? (G2)
C	10	13	11 th	The Treaty of Versailles	-
	11	14	10 th	Crusades	-
	12	15	11 th	The time of regents and monarchs	-
		16	10 th	The time of monks and knights	-

Table 4.3. Coding Scheme for Teacher-Developed Tasks

Design Principles	Task ID													
	A1	A2	B1	B2	C	D1	D2	E1	E2	F1	F2	G1	G2	
<i>#1 Evaluative question</i>														
Present yes/no	yes	yes	yes	yes	yes*	yes	yes	yes	yes	yes	yes	yes	yes	yes
<i>#2 Multiple sources</i>														
Present yes/no	yes	yes	yes	yes	yes*	yes	yes	yes	yes	yes	yes	yes	yes	yes
Number of sources	6	7	7	5		5	6	2	4	5	5	4	5	
Number of textual sources	3	5	7	5		4	6	2	3	4	4	2	3	
Number of visual sources ¹	3	2	0	0		1	0	0	1	1	1	2	2	
Mean length of source texts	201	183	211	168		200	162	455	312	175	146	131	105	
<i>#3 Fit</i>														
Students' grade level	11 th	11 th	10 th	10 th	12 th	11 th	11 th	11 th	11 th	11 th	11 th	11 th	11 th	
Difficulty as experienced by teachers ²	3	2	2	2	2	2	2	2	2	2	2	2	2	
<i>#4 Task duration</i>														
Appropriate for one lesson hour ³	2	2	3	3	3	2	2	1	1	2	2	2	2	
Duration of lessons	50'	50'	45'	45'	60'	50'	50'	50'	50'	50'	50'	40'	40'	

¹ Pictures, tables, cartoons, or videos.

² Difficulty: 1 = easy, 2 = not too difficult, not too easy, 3 = difficult.

³ Task duration: 1 = too short for one lesson hour, 2 = not too short, not too long, 3 = too long for one lesson hour.

* Students formulated their own evaluative question and searched for appropriate source materials themselves.

2.3.2 *Description of the Intervention*

In Condition WT+S, strategy-instruction as well as evaluative writing tasks were implemented. In Condition WT, such evaluative writing tasks were implemented without strategy instruction. The interventions as designed for Condition WT+S and WT are presented in Table 4.4.

The strategy instruction for Condition WT+S consisted of two lessons. The instructional design (how to teach the strategy) was roughly based on Harris and Graham's classic SRSD-framework (1996), and consisted of five instructional stages: (1) development of background knowledge, (2) describe it, (3) model it, (4) support it, and (5) independent performance.

In Lesson 1, direct strategy instruction was provided. Central to this was a history-specific strategy for performing evaluative writing tasks: the Read-Think-Write strategy (see Chapter 3). This RTW strategy, is a seven-step strategy directing students how to read source texts diligently, how to connect information from different sources, and how to write a text with the reader in mind. In the second instructional stage, model it, it was advised to show students how a strategy is used. In an attempt to unburden the teachers - who might feel challenged by having to model unfamiliar reading and writing processes - a video with students modeling the RTW strategy was provided for usage. Each teacher could decide to simply show the video on the smart board, and discuss it in a class discussion.

In Lesson 2, students read three example texts, and discussed them with peers. In a class discussion, a list of criteria for good evaluative history texts was constructed.

After the two strategy lessons, students from Condition WT+S performed the writing tasks as developed by their own teacher. Each task was discussed in class in a feedback session, with the feedback focusing on the writing process. In Condition WT, students performed two writing tasks, as developed by their own teacher, and these writing tasks were discussed in feedback lessons with product feedback, focused on content. For both conditions, students were allowed to collaborate in the prewriting phase of Lesson 3 and 5 (Table 4.4), but they wrote their texts individually. Writing tasks were all performed on a computer, using laptops in the classroom, or computers in a computer room.

Table 4.4. Description of the Intervention

Lesson	Stage	Learning activities	Condition WT+S	WT
0	<i>Develop background knowledge</i>	Task experience	Students perform a first evaluative writing task on an uninstructed historical issue (= pre-test).	A
1		Reflection	The students write down what they thought was easy while performing the task from Lesson 0, and what was difficult.	-
	<i>Describe it</i>	Building new knowledge	The teacher presents and explains the <i>Read-Think-Write</i> strategy.	
		Comparing strategy to own experience	The students compare their own experiences with evaluative tasks to the presented strategy.	
	<i>Model it (process)</i>	Observing strategy demonstrated by a modeling peer (video)	Students watch a 12-minute video in a plenary session on the main screen. This video presents the strategy and contains fragments showing modeling peers, who demonstrate how each step of the strategy could be performed.	
		Assessing a peer's performance	As a processing activity, the students individually compare the performance of the modeling peer with the strategy as presented by scoring the peer on a scale from 0 to 100.	
2	<i>Model it (product)</i>	Assessing peers' texts	Students individually assess three example texts.	-
		Generating criteria	In a class discussion, students generate a criteria list.	
		Applying new learning	Students apply the criteria; they revise the text they wrote (Lesson 0), with the criteria list in mind.	
3	<i>Support it</i>	Scaffolded practice	Students perform a next evaluative writing task, scaffolded by the support of peers and the teacher.	A
			Written guide available.	B
4	<i>Support it</i>	Teacher/Peer feedback	The teacher chooses how to provide feedback from the suggestions in the teacher manual.	A
			Feedback is focused: on <i>process</i> .	C
5	<i>Independent performance</i>	Individual work with the help of a written guide	Students perform a third evaluative task.	A
			Written guide available.	B
6	<i>Support it</i>	Teacher/Peer feedback	The teacher chooses how to provide feedback from the suggestions in the teacher manual.	A
			Feedback is focused: on <i>process</i> .	C
7	<i>Transfer</i>	Individual work	Students perform a fourth evaluative source-based writing task on an uninstructed historical issue (post-test).	A

Note: A = As WT+S; B = No guide; C = Feedback is focused on content.

As presented in Table 4.4, Lessons 4 and 6 centered on feedback on students' texts. The teachers were provided with condition specific suggestions on how to provide this feedback. These suggestions can be found in the supplementary materials. The feedback lessons were designed to be low cost for teachers, since it is generally known that the workload of providing feedback on written products is one of the main reasons for teachers to avoid writing tasks as learning activities. In the teaching manual for Condition WT+S, the suggestions all encouraged the teachers to focus their feedback on students' writing process. In the teaching manual for Condition WT, teachers were guided towards providing product feedback focused on content.

2.3.3 Control Condition

Between the pre- and the post-test, students in the Control Condition attended regular history classes, as they were used to. This mostly involved the teacher explaining topics within the classroom, and students working on assignments from the workbook, individually or in groups. Students wrote notes and short answers to questions, but no larger writing assignments were performed.

2.3.4 Implementation Fidelity

To check whether the intervention was implemented as intended, we included triangulated measures: the research team collected teacher logs, did observations, and conducted reflective interviews with both students and teachers. Teachers from all three conditions kept a guided logbook during the intervention period. In Conditions WT+S and WT, LH observed one intervention lesson for each group. After the intervention period, LH interviewed each teacher of Condition WT+S and WT, as well as a group of three to four students from each group within these conditions. Interview guides can be found in Appendix E.

The teachers were free to plan their lessons as they preferred, resulting in different durations of the intervention. The mean duration between pre- and post-test was 8.3 weeks ($sd = 3.2$ weeks). A consequence of the free planning was that the number of history content lessons also differed between groups and conditions.

Concerning the intervention lessons, the reflective interviews and observations exposed several deviations from the intervention-as-designed in Groups 5 and 7. In Group 5, there was only one extensive writing task instead of two (as previously discussed). Since this task was more time-consuming, the order of activities was adjusted according to this. The task was explained

at the beginning of the intervention, as students needed time to think of an evaluative question, and to search for source materials. The students wrote their texts after Lesson 1, they then discussed example texts in Lesson 2, and students were given the opportunity to revise their respective texts based on this discussion, prior to handing in their work for a formal assessment. The students all received individual feedback, as the task was a formal exam, and thus there were no feedback lessons.

Group 7 deviated from the design as constructed, due to presumed lack of time: the students performed the two evaluative tasks (Lesson 3 and 5) both in one 50-minute session. Both tasks were discussed in the subsequent lesson in one session as well, thus, Lesson 4 and 6 were also merged.

Furthermore, Groups 8 and 12 were not provided with sufficient time for the pre-test. Their intervention-as-implemented was also different from intended; the students worked individually, and on their own pace, through the materials. Therefore, these two groups were excluded from further analyses.

Regarding the options for providing feedback, the options each teacher chose for the two feedback lessons were coded (Table 4.5). Next to the five suggestions presented in the teaching manual (coded #1-#5), there were two more codes. Code #6 resembled personal feedback. Some teachers chose to provide students with personalized written feedback, despite our discouragement, due to lack of lesson time to discuss example texts during lesson hours. In some cases, the teachers graded students' texts, and therefore chose to provide written feedback as well. Code #7 resembled only a general discussion of content. For instance, some teachers discussed which perspectives should be mentioned in the text, in order to discuss the issue properly. The teachers who gave feedback coded as #7, explained this was due to time constraints.

As can be appreciated in Table 4.5, three teachers chose to discuss mainly content after the second writing task. However, they had already discussed the first task extensively.

Thus, all teachers except for Teacher 7 (Groups 8 and 12) implemented Lesson 1 to 5 of the intervention as intended, as all essential elements were implemented, and this implementation was of sufficient quality. Lesson 6, however, was not considered optimal by three of the teachers. These teachers explained that lack of time played an important role. However, when a teacher is confronted with time constraints, this means they are forced to prioritize. Priorities were apparently not set on providing feedback on students' texts, but more on knowledge accretion. For example, teacher 9 explained that they had read students' texts for task F2, and then decided to skip lesson 5:

Teacher 9: "I think I was shocked at how they were lacking in content in their texts, and then I thought: I have to focus on that now, because otherwise their final test results will be a complete disaster."

Nevertheless, it was decided to include the results of these three teachers as well. Feedback was of lesser quality after the second task, but not absent in their intervention as a whole.

Table 4.5. Coding of Feedback Lessons

Condition	Teacher	Group	Feedback Suggestion				
			1	4	5	6	7
WT+S	1	1			A1	A2*	
	2	2		A1		A2*	
	3	3&4	B1			B2	
	4	5				C*	
	5	6	D1				D2
	6	7	E1&E2				
WT	8	9&10	F1				F2
	9	11	G1				G2

Notes: Feedback suggestions: (1) creating a scale; (2) comparing text to a scale (not used); (3) discussion of random text (not used); (4) discussion of good example; (5) assessing on criteria; (6) personal feedback; (7) discussion with focus on content only. For the task-ID's, see Table 4.2.

*Texts were graded by the teacher.

2.4 Variables and Measurements

To answer the research questions, the effects of the intervention on three aspects were measured. Firstly, the students' historical writing proficiency (RQ1-2a) was measured, using source-based writing tasks (pre-test/post-test). Secondly, the epistemic experience (RQ1-2b) was measured, using a statement to rate about the epistemic experience immediately after performance the writing task (pre-test/post-test). Thirdly, course content knowledge (RQ1-2c) was measured, using a recall test (pre-test/post-test). Furthermore, to answer RQ3, all students completed two questionnaires before the start of the intervention.

2.4.1 Writing Proficiency

To measure source-based writing proficiency (RQ1-2a), students performed writing tasks about history topics which were unrelated to course-content. The post-task might thus be considered a transfer task. In the pre-test, the main question was: To what extent do you think Western museums should

return colonial art to the country of origin? Four source texts were provided ($\mu = 218$ words). For the post-test, the task was about the Dutch police actions in Indonesia (1947-1949)*. The main question here was: To what extent can the term "police action" be justified, in hindsight? Again, four source texts were provided ($\mu = 198$ words). Both tasks contained mostly primary sources, representing multiple perspectives. Each task took about 30 minutes to perform.

We controlled for condition effects on task perception on four dimensions: prior knowledge, topic interest, effort and difficulty. This was done by asking all students to rate four statements (Table 4.6), on a scale of 1 (completely disagree) to 5 (completely agree), in order to explore how each student had perceived their tasks.

Table 4.6. Task Perception Measurement

Prior knowledge	<i>I knew a lot about the topic already before I'd read the source texts.</i>
Topic interest	<i>I think the topic of this task is interesting</i>
Effort	<i>I put a lot of effort into this writing task.</i>
Difficulty	<i>I thought this writing task was difficult.</i>

2.4.2 Epistemic Experience

A measurement to obtain insight into the students' knowledge gain of the writing task topic, was incorporated into a question at the very end of the pre-/post-test on epistemic experience (RQ1-2b). We asked students to rate the statement "Through this writing task I learned a lot about the topic" on a scale of 1-5. "The topic" referred to the topic of the respective source-based writing task. We asked this question to explore if the students felt they had gained knowledge on a topic, unrelated to the lesson content, through the task set.

2.4.3 Course Content Knowledge

To measure the effect of the replacement of textbook tasks by writing tasks, and the addition of strategy instruction, on course content knowledge (RQ1-2c), we administered a ten-minute open recall test, as the first measurement in the pre-test and the post-test sessions (Casado-Ledesma, et al., 2020; Langer & Applebee, 1984). The question used was: "What do you know about ...". On the dots, each teacher filled out the topic they discussed in the upcoming (pre-

*In 1947, the Dutch government did not recognize the Republic of Indonesia as an independent state, but regarded it as a rebellious movement within the colony of the Dutch East Indies, against which "police action" had to be taken.

test) or past (post-test) history lessons, during the intervention period. These topics differed per group (Table 4.1).

2.4.4 Learner Variables

We administered two questionnaires before the pre-test session: a writing-beliefs questionnaire, and a self-efficacy questionnaire (RQ3). The teachers instructed their students to complete these in a lesson hour prior to the start of the intervention. It took students about 10 minutes to complete both questionnaires, using Qualtrics. Both questionnaires can be found in Appendix C.

The writing-beliefs questionnaire consisted of 26 statements about writing in general. We distinguished four aspects (transmission; high amount of revision; emotional engagement; cognitive engagement), based on the validated questionnaire of Vandermeulen (2020). All students were asked to indicate to what extent they thought each statement was true, on a scale from 1 (totally disagree) to 5 (totally agree).

The self-efficacy questionnaire was developed and tested in a former study (Chapter 3). It consisted of 30 'I can'-statements, which were all related to aspects of the historical reading-writing process. All students were asked to indicate, to what extent they thought the statements were true, on a scale from 0 to 100. The consistency of the questionnaire was high (Cronbach's alpha .93). Although a factor analysis uncovered the structure of six factors (for details, see Appendix F), we maintained students' total self-efficacy scores in our analyses, based on our exploratory goal.

2.5 Data Preparation

2.5.1 Writing Proficiency: Text Rating Procedures

Students' texts were assessed in two ways: analytically, using a rubric with five dimensions, and holistically, using a scale with annotated anchor texts. With the rubric, assessors were asked to score the students' texts on five dimensions: situational understanding, multiperspectivity, argumentation, source use, and structure (derived from De La Paz et al., 2017). Subsequently, they scored each text holistically, with the help of a scale. This scale was developed by the research team for a previous study (presented in Chapter 3), and contained five anchor texts, representing scores 70-85-100-115-130. The anchor texts in the scale were accompanied by annotations, pointing towards five foci: the dimensions of the rubric. It resulted in six scores for each text. Lastly, a final indicator for text quality was text length (number of words).

The data set consisted of 277 texts. Per topic, the texts were randomly divided over sets containing 8-10 texts. We excluded 11 texts (pre-test: 5; post-test: 6) which were unfinished (containing only a few words accompanied by a remark such as 'I had no time left'), or were a reflection of the task not being fully understood (for example when students noted: 'I have no idea what to do with this task' or 'I don't understand the source texts'). Of these excluded texts, nine were from the control condition group.

Because the data set consisted of a large number of texts, the assessments were performed by three jury panels. In total, 21 history teachers (e.g., former teachers, teacher trainers, teacher students) as well as LH, took position in the jury panels. A system was set-up, in order to have each jury member rate six sets (= approximately 60 texts in total, 30 texts from the pre-test, 30 texts from the post-test), while (partly) overlapping with four other raters. This way, each text was assessed by three jury members. Intraclass correlation coefficients of the juries were between .79 (pre-test argumentation) and .89 (post-test structure) ($\mu = .84$), which was considered satisfactory, and further analyses were consistently conducted with the mean of the three jury's scores.

2.4.2 Course Content Knowledge: Recall Test Analyses

It was hypothesized that students with more knowledge-gain, would write more new information during the post-test. Therefore, each student's pre- and post-recall tests were compared, and the new T-units (minimal terminable units (Hunt, 1965)) were counted. For example, if a student wrote down three T-units in the pre-test and three in the post-test, the number of T-units from the post-test that contained new or extended information compared to the pre-test were recorded. When there were three new information units in the post-test, the student was given the score '3'. T-units had to be correct, and relevant, to be counted as 'new'. A subset (10%) was analyzed by a second rater, in order to check the reliability of scoring (ICC = .95).

Secondly, the mean length of new T-units was calculated, since it was hypothesized that students in the writing conditions would write richer, and more new units containing connectors and adverbs, indicating more relations between concepts. Generally, relations within units are associated with more elaborated knowledge, and understanding of content.

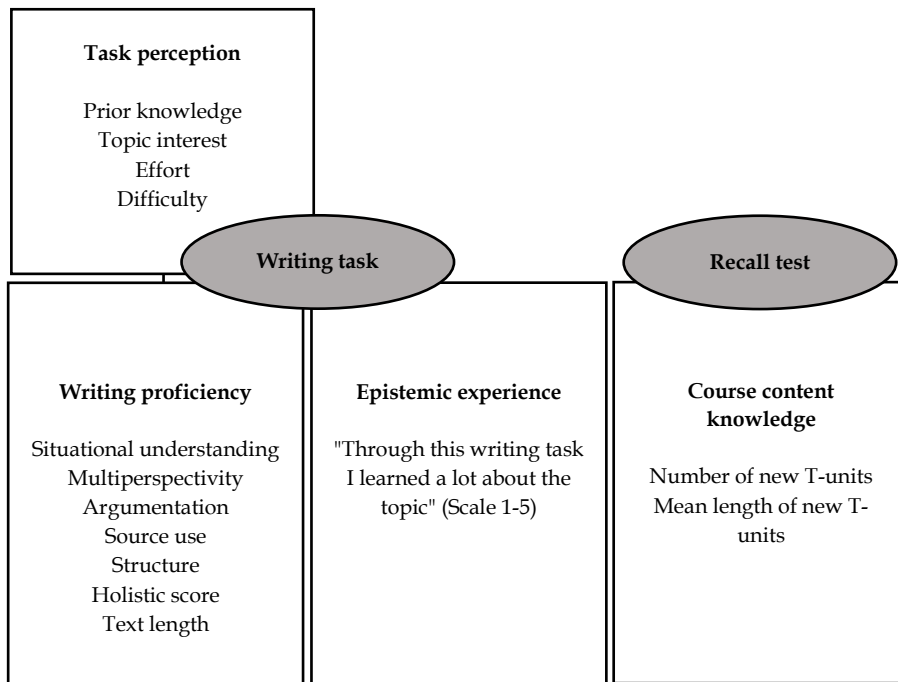
Those students' recall tests which were completely blank (i.e., 2.3%) were excluded from the analyses, as the underlying reasons were unknown. Conversely, responses only consisting of remarks, such as 'I don't know anything

about this', were included: wit the word/T-unit count for tests only containing such a remark being 0.

2.5 Data Analysis

We conducted a multilevel analysis for three dependent variables: (1) writing proficiency, (2) epistemic experience, and (3) course content knowledge. An overview of variables is presented in Figure 4.1. The seven indicators for text quality were closely related (r between .77 and .90). The holistic score correlated best with the other six indicators (r between .83 and .90), based upon it was decided to focus the analyses on this score, when analysing writing proficiency.

Figure 4.1. Overview of Variables



The effects of the interventions were tested by comparing four nested models, with Group as random factor, and time as repeated measure, including Group*Subject as random factor: Model 0 with random factors only, Model 1 with the added factor Time, Model 2 with the added factor Condition, and Model 3 with the added interaction Time*Condition. To gain insight into the effects of the Condition at the post-test measurement, the analysis for the best fitting model was reran, using dummy variables for the effect of Condition (three levels), and Time (two levels), and the resulting pair-wise comparisons were checked.

For the course-content knowledge indicators (number of new T-units and length of new T-units), two models were compared, while Time was not a factor: Model 0 with Subject as random factor, Model 1 with added factor Condition.

To explore moderating effects of self-efficacy and beliefs, we centered the scores, and expanded the models presented above, with Model 4 (general effect of the learner variable LV); Model 5 (interaction LV with Time); Model 6 (interaction LV with Condition); and Model 7 (interaction LV with Time and Condition). We ran these series of models for holistic score (for writing proficiency), and epistemic experience. All model comparisons can be found in Appendix G.

2.6 Preliminary Analyses

2.6.1 Pre-test Differences

We conducted preliminary analyses, to explore differences between conditions, on all outcome variables (all writing proficiency variables, epistemic experience, and all three variables for course content knowledge), at the pre-test. For none of the dependent variables, the learner variables, and task perception variables, effects of condition were observed at the pre-test.

2.6.2 Task Perception

Because the pre-test task was an essential part of the intervention lesson (Lesson 0), we could not randomly assign pre- and post-test topics. Topics were thus nested within measurement occasions. Therefore, we checked for pre-test/post-test similarity, on four aspects: prior knowledge, topic interest, effort, and difficulty. Mean scores for the four statements on task perception, are presented in Table 4.7.

A multilevel analysis revealed effects of the pre-test/post-test writing task. We found differences on two out of four aspects. Students felt they had more

prior knowledge in the post-test than in the pre-test ($F(1,223) = 9.73, p = .002$), and they perceived the post-test as more difficult ($F(1,235) = 22.07, p < .001$). Additionally, the students perceived the two tests as similarly interesting, with a relatively high mean score, indicating the tasks were indeed experienced as quite interesting. The students also indicated that they had put similar effort into their performances of the either of the two tasks. We found no interaction effects with condition, meaning that possible differences between conditions on outcome variables were not influenced by differences in these task-perception variables.

Table 4.7. Task Perception Scores (Scale 1-5)

	Condition WT+S		Condition WT		Control Condition	
	Pretest <i>M (SD)</i> <i>n</i> = 107	Posttest <i>M (SD)</i> <i>n</i> = 91	Pretest <i>M (SD)</i> <i>n</i> = 55	Posttest <i>M (SD)</i> <i>n</i> = 55	Pretest <i>M (SD)</i> <i>n</i> = 70	Posttest <i>M (SD)</i> <i>n</i> = 78
Prior knowledge	2.58 (.98)	2.71 (1.04)	2.78 (.92)	3.02 (1.05)	2.39 (.90)	2.80 (1.10)
Topic interest	3.54 (.92)	3.64 (.89)	3.67 (.90)	3.82 (.82)	3.44 (.92)	3.51 (1.00)
Effort	3.38 (.79)	3.45 (.73)	3.19 (.58)	3.16 (.71)	3.07 (.78)	3.30 (.94)
Difficulty	2.74 (.91)	3.00 (.90)	2.59 (.84)	3.06 (.91)	2.71 (.99)	3.14 (1.04)

2.6.3 Learner Variables

The students' mean scores for learner variables are presented in Table 4.8. For self-efficacy, the mean individual total score was 66.2 ($N = 242$, $\min = 39.7$, $\max = 95.5$, $sd = 10$). For beliefs, the scale "revision" was not reliable ($\alpha = .47$), similarly to in the studies of Vandermeulen (2020). Based on this, no further analyses with this scale were conducted. The three remaining beliefs scales did not correlate highly (varying from $-.031$ to $.314$), which lead us to consider them separate. There were no significant differences between conditions.

Table 4.8. Mean Scores for Beliefs (Scale 1-5) and Self-Efficacy (Scale 0-100)

	Condition WT+S <i>n</i> = 115 <i>M (SD)</i>	Condition WT <i>n</i> = 61 <i>M (SD)</i>	Control Condition <i>n</i> = 78 <i>M (SD)</i>
Transmission (7 items, $\alpha = .67$)	2.88 (.56)	2.92 (.52)	2.90 (.51)
Emotional Engagement (5 items, $\alpha = .59$)	3.36 (.53)	3.35 (.56)	3.33 (.48)
Cognitive Engagement (3 items, $\alpha = .77$)	3.65 (.77)	3.80 (.63)	3.55 (.80)
Self-efficacy	65.6 (9.5)	68.9 (10.6)	65.2 (10.1)

3. RESULTS

In the following section the intervention effects on students' writing proficiency, epistemic experience, and course content knowledge are presented. The focus lays on the differences between conditions, as pre-test differences were checked for with preliminary analyses.

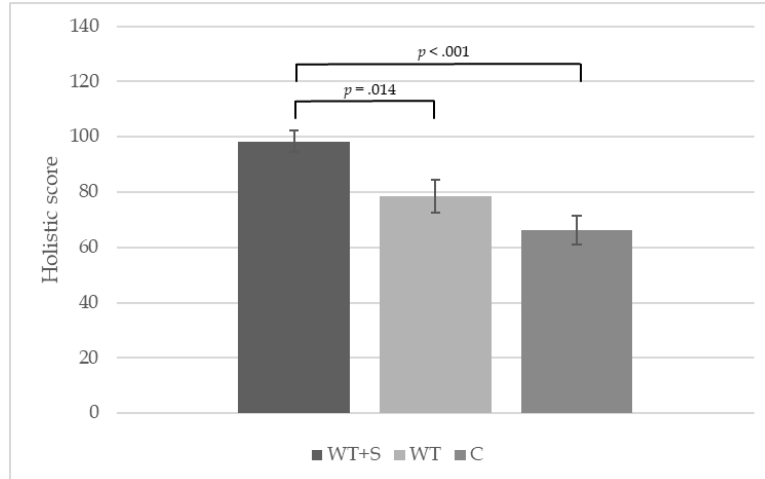
3.1 Writing Proficiency (RQ1-2a)

For holistic score, Model 3 fitted the data best, as no differences were observed at the pre-test. This revealed an interaction effect of Time and Condition, and therefore an effect of conditions on the post-test (Figure 4.2).

In the post-test, the WT+S Condition scored significantly higher than the WT Condition (mean difference: 19.68, $se = 7.1$, $p = .014$), and than the Control-condition (mean difference: 32.02, $se = 6.47$, $p < .001$). Condition WT did not differ significantly from the Control Condition with regard to holistic scores ($p > .05$).

Moreover, similar results were observed for all other indicators of text quality: situational understanding, multiperspectivity, argumentation, source use, structure, and text length.

Figure 4.2. Holistic Scores in the Post-test, as Estimated Under Model 3

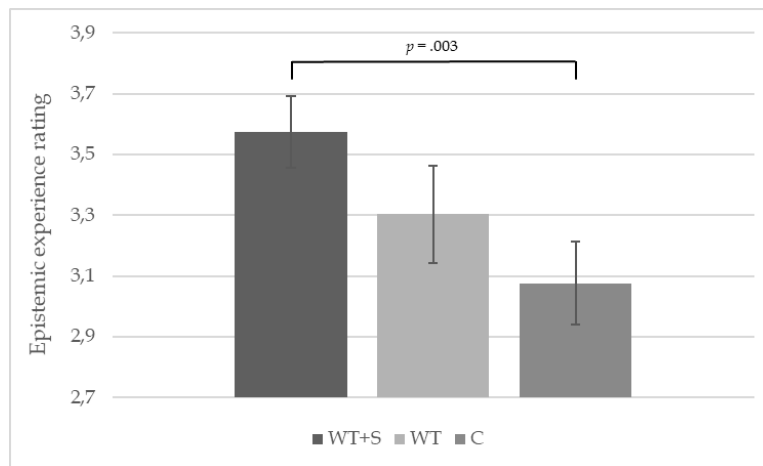


3.2 Epistemic Experience (RQ1-2b)

For epistemic experience, it was also Model 3 which fitted the data best, suggesting an interaction effect of Time and Condition (Figure 4.3). In the post-test, students in Condition WT+S rated their epistemic experience

significantly higher than the those in the Control Condition (mean difference: .492, $se = .149$, $p = .003$).

Figure 4.3. Epistemic Experience Ratings (Scale 1-5) in the Posttest, as Estimated Under Model 3



3.3 Effects on Course Content Knowledge (RQ1-2c)

For the two indicators of course-content knowledge, Model 1 fitted best, indicating an effect of Condition. Results are presented in Table 4.9. For number of new T-units, condition WT scored significantly lower than Condition WT+S (mean difference: -1.80, $se = .72$, $p = .013$) and than the Control Condition (mean difference: -2.11, $se = .77$, $p = .006$).

For mean length of new T-units, however, results were different. While the Control Condition scored significantly lower than Condition WT (mean difference: -2.62, $se = .97$, $p = .007$), there was no significant difference for Condition WT+S with the other conditions.

In summary, although students in Condition WT seemed to write less new T-units, their new T-units were relatively longer. While those in the Control Condition did write significantly more new T-units; those new T-units were relatively shorter.

Table 4.9. Estimated Means for Course Content Knowledge Indicators, Model 1

	Condition WT+S EM (SE)	Condition WT EM (SE)	Control Condition EM (SE)
Number of new T-units	6.41 (.43)	4.61 (.58)	6.72 (.50)
Mean length of new T-units	9.34 (.54)	10.35 (.73)	7.73 (.63)

3.4 Moderating Effects of Learner Variables (RQ3)

We tested whether learner variables moderated the intervention effect. For writing proficiency, there were no moderating effects of learner-variables. We did observe a better fit in Model 4, with the addition of cognitive engagement and self-efficacy (Appendix G), indicating a direct effect of those factors on writing-proficiency in all conditions. The effects were positive: students with higher self-efficacy scores, and students reporting more cognitive engagement beliefs, obtained higher holistic scores (Table 4.10).

Table 4.10. Estimates of Fixed Effects on Holistic Score for Two Covariates, Model 4

Moderators	<i>b</i>	<i>SE b</i>	95% confidence interval		<i>p</i>
			lower bound	upper bound	
Cognitive engagement	3.88	1.43	1.06	6.70	.007
Self-efficacy	.22	.11	.002	.44	.048

Furthermore, we have observed a moderating effect of emotional engagement and of self-efficacy, on *epistemic experience* (Appendix G, Model 7). In Condition WT and the Control Condition, these learner variables seemed to have no effect, while in Condition WT+S the two moderators did seem to contribute to the effect.

In condition WT+S, epistemic experience ratings developed differently for the different levels of emotional engagement beliefs: they remained unchanged from pre- to post-test for low-level emotional engagement beliefs, while they increased for high-level emotional engagement beliefs ($b = -.86$; $se = .32$, $p = .007$) (Figure 4.4). The same pattern was observed for self-efficacy (Figure 4.5). These results indicate, that the strategy instruction seemed more effective for students with high self-efficacy or high emotional beliefs levels, similarly to the aforementioned epistemic experience.

Figure 4.4. Epistemic Experience in Condition WT+S versus Control Condition for Students With Different Levels of Emotional Beliefs (= EB)

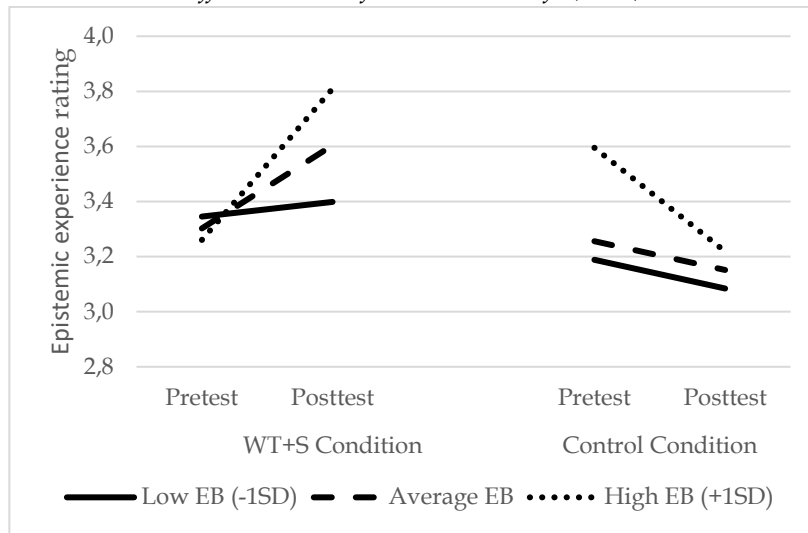
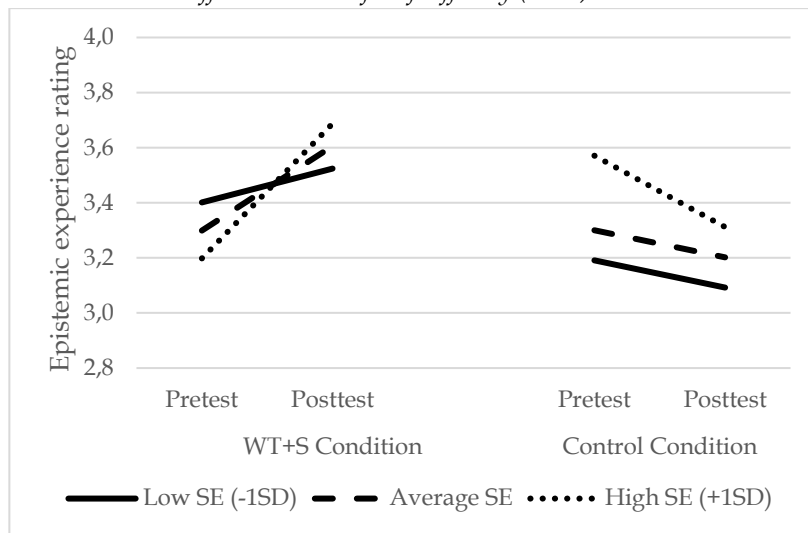


Figure 4.5. Epistemic Experience Condition WT+S versus Control Condition for Students With Different Levels of Self-Efficacy (= SE)



4. DISCUSSION AND CONCLUSIONS

In the current study, we aimed to investigate whether integrating writing tasks as learning activities into the history classroom, when accompanied by strategy instruction, would foster students' historical writing and history learning (pre-university education, grades 10-12). In order to do this, we proposed three research questions. Specifically, we have investigated whether integrating writing tasks as learning activities into the history classroom (RQ1), and additional discipline specific writing strategy instruction (RQ2), would improve the students' disciplinary writing proficiency, epistemic experience, and course content knowledge. Additionally, we have explored if certain learner variables moderated intervention effects (RQ3).

With regard to RQ1, we hypothesized that the replacement of workbook assignments by source-based writing assignments, would lead to greater epistemic experience in a transfer task, and more course-content knowledge in a recall task, but not to more advanced writing proficiency. Our results confirmed our hypotheses only in part. From the data, we can conclude that only replacing text book assignment by writing tasks (Condition WT), was indeed not sufficient for students to significantly outperform a non-writing control group on writing proficiency. These results confirm Monte-Sano's previous finding (2008), being that the act of writing alone was not sufficient for students to improve their evidence-based history essay writing. Furthermore, Condition WT did not outperform the Control Condition when assessing the epistemic experience in the transfer writing task, which was unlike our expectation. We had expected the students in the writing conditions, to experience the epistemic function of writing, especially after performing writing-to-learn tasks. Upon assessing course-content knowledge, students in the WT Condition wrote less new T-units in a post-test recall task, compared to the Control Condition, however, their T-units were significantly longer. Since longer T-units are associated with more clauses, implying more connections between elements, it could be conceivably concluded that the WT Condition was at least not inferior to the Control Condition with regard to course content knowledge. This finding seems in line with Klein's hypothesis (1999), that writing holds the intrinsic value of learning: writing in itself, is an instructive activity.

Concerning RQ2, we had predicted that those students who received additional writing-strategy instruction, would achieve higher text-quality scores than those students who performed writing tasks without process instruction, or than those students in a non-writing condition. We expected the strategy

instruction to reinforce the effects on epistemic experience, and course-content knowledge. From the data obtained, a possible conclusion could be that the inclusion of a discipline specific strategy instruction indeed effectuates an improvement of students' writing proficiency. Specifically, for Condition WT+S, students outperformed Condition WT and the Control Condition on *all* aspects of text quality: holistic score, text length, and discipline-specific criteria – situational understanding, multi-perspectivity, argumentation, source use, and structure. These results are similar to previous results of strategy instruction (e.g. De La Paz et al, 2017). Furthermore, Condition WT+S also scored the highest on epistemic experience.

With regard to course content knowledge, there were no differences observed between Condition WT+S and the Control Condition. The inclusion of strategy instruction and writing tasks, thus seems not to impact the course content knowledge, while, at the same time, it does seem a better option for obtaining disciplinary writing proficiency, and for a better epistemic experience, than merely implementing writing tasks.

When we include learner variables into the interpretation of the results (RQ3), two sets of results were observed: general and moderator effects. We found direct effects of cognitive engagement-beliefs and of self-efficacy, on holistic text quality; meaning, that students who believed that writing and thinking are intertwined, wrote better texts. A similar conclusion was drawn by other researchers (Villalón et al., 2015). A possible explanation for this, could be that students who consider writing tasks as learning tasks, perceive writing as more useful. However, we could not confirm this with the data; there were no significant correlations between cognitive engagement beliefs, and reported task effort. The general effect of self-efficacy on text quality, does confirm previous research outcomes (e.g. Sanders-Reio et al., 2014).

Furthermore, we found two moderator effects for epistemic experience in Condition WT+S only: emotional engagement and self-efficacy. Students with high emotional engagement beliefs and/or with high self-efficacy, seemed to benefit most from the strategy instruction; they showed more epistemic experience.

Although history teachers have generally tended to underrate writing process instruction, and might not consider this a priority, our results indicate that if teachers want to teach students to write historically, writing opportunities with writing process instruction is vital. The present study has shown that teachers can replace workbook questions on any topic, with evaluative writing tasks, and add strategy instruction. These additions not only increase

historical writing proficiency in transfer tasks, but also stimulate to experience the process of writing as epistemic, and expand students' course content knowledge. This is particularly important, as from a history teachers' perspective, a lagging growth of content-knowledge would be problematic. Providing process instruction is already often seen as a 'waste of time' for building knowledge, which would be even more so, if recall tests had confirmed this. However, our results have shown that writing tasks as learning activities were not harmful for course content knowledge at all. In fact, the recall test results showed that writing might even provoke more complex knowledge, as evidenced by the relatively extensive length of new units. Moreover, the results even suggest that students improved their historical reasoning, considering the higher scores on situational understanding, argumentation, and source use, which might be seen as indicators for this skill (Van Boxtel & Van Drie, 2018; Van Drie et al., 2014).

4.1 Strengths, Limitations, and Future Research

In this study, we aimed to overcome known obstacles for teaching literacy, such as the persistent 'pedagogy of telling'. The inclusion of writing-to-learn tasks was perceived as successful: the teachers of Conditions WT+S and WT experienced the writing tasks as instructive learning activities. During the evaluative interviews, the teachers indicated that they considered it 'the essence of history' to learn to think about complex historical issues from several perspectives. This makes writing evaluative texts a valuable activity for learning history, in a different way than the text book dictates.

Another obstacle to overcome, was teachers' lack of knowledge about writing processes. We did this, by using a solid instructional design with feasible tools (e.g., a video with modeling students), and by offering a PD-session in order to secure the correct and effective use of the intervention materials.

An important feature of the design of this study, was the flexibility of our intervention, in order to make the design feasible, and easy to implement, into the curriculum. We provided the teachers with many possibilities: teachers were not bound to a specific topic, a specific grade level, or to a narrow time-frame. Moreover, the students were not constrained by a fixed strategy either: we provided different routes, in order to accommodate each student's own writing routines. This research design also increased generalizability; the effects are not bound to a certain course within history education, as various thematic units were involved, in all conditions. Future research could focus

on the effectiveness of similar disciplinary approaches in other school subjects.

Although we consider adaptability an important strength of our instructions, it might also be seen as a limitation. The differences between groups, and thus also between conditions, may have resulted in different outcomes. However, we tried to maintain consistency between groups, and between conditions, during the intervention, by a) establishing and monitoring clear design principles for writing tasks and strategy instruction; b) using statistical analyses which level out group differences; and c) measurements, for example, the recall test contained a pre- and post-test; controlling for prior knowledge of the different topics. However, despite these precautions, groups remained unequal, for example with regard to the amount of time spent on content-knowledge building. Also, the groups differed with respect to duration of the intervention, implying different numbers of content lessons. Although this latter was perhaps to be expected, as topics differed from group to group, and additionally, topics were not equally difficult or complex.

Furthermore, the teachers' positive evaluations of the writing tasks, attest that they prefer discussing historical issues in depth over discussing many different topics superficially. In the writing conditions WT+S and WT, the topic has to be discussed in depth, since writing tasks provoke this. However, the approach of covering a large set of topics on a very shallow level ("a mile broad, an inch deep"), is quite common in history education, and this thus might have been the dominant approach in the control condition. This could be seen as a limitation, since there is no clear view on the interference between complexity of the topic and time spent. Moreover, we could question the feasibility of meeting all curriculum demands with this in-depth approach, in the long term. This concern could, in turn, restrain teachers from applying writing tasks more widely.

Another possible limitation is that the recall test was used as a surrogate for course content knowledge. Previous research has criticized recall tests as a measure of knowledge gain through writing, since this test is likely to detect knowledge accretion, which is only a peripheral outcome of writing (Schumacher & Nash, 1991). However, the recall measure was not the only content measure in our study; in the assessment of the students' texts, we deliberately included content-related indicators, such as argumentation, as well.

A final limitation worth mentioning, is that the data from the current study, is inconclusive about the potential durability of the effect observed. In

short, it has not been made clear if a writing task later on in the school year, would need to be accompanied by another process instruction, or if a quick reminder of the strategy would be sufficient. This question of durability of the process regarding instruction for students' disciplinary literacy, should be further examined in future research.

4.2 Conclusions

In conclusion, after only two strategy lessons and two writing opportunities for practice, of about 50 minutes each, students overall wrote better texts. Compared to the WT Condition, where students had similar writing opportunities, but were uninstructed, the added process instruction made a significant difference. Moreover, in the setting in which there were writing opportunities, but without the instructions, outcomes were comparable to those for students in a control condition without writing. Although history teachers generally tend to underrate writing process instruction, and do not consider it a priority, our results underline the necessity, and the importance, of including writing strategy instruction into their discipline anyway; and contradict the archaic paradigm, that including writing tasks and process instruction are a 'waste of time'.