

WRITING TO LEARN HISTORY:

An instructional design study

LIEKE HOLDINGA, JANNET VAN DRIE, TANJA JANSSEN,
& GERT RIJLAARSDAM

Research Institute of Child Development and Education, University of Amsterdam

Abstract

This study reports on the design and evaluation of an instructional unit, aimed at improving secondary school students' disciplinary writing in history. Central to this design was the replacement of conventional workbook exercises by evaluative source-based writing tasks which were co-developed with participating history teachers. Additionally, an instructional unit to teach students a discipline-specific reading-thinking-writing strategy based on previous research was designed. Two history teachers implemented the evaluative tasks and the strategy instruction in their 11th grade history classrooms in a trial intervention study with a switching panels design. Pre-, mid-, and post-testing consisted of evaluative writing tasks (ca. 200-300 words), which were analyzed on holistic quality, content quality, quality of structure, and text length. Results showed effects in the second panel for content quality.

In this paper we elaborate on the design of this strategy and the instructional design, as well as the design principles underpinning these. Based on the trial study, we present recommendations for redesign in order to optimize practicality and effectiveness of the instructional unit.

Keywords: disciplinary literacy, historical writing, writing education, instructional design, design study

1. INTRODUCTION

Subject areas in upper secondary education demand a high level of literacy. Students are presumed to be able to read, and write, complex texts which are tuned to the relevant discipline. Improving literacy should thus play an important part in education; this insight has been shared by educational researchers for many years now. Around 1980, a general approach to teaching literacy was common: it emphasized the value of literacy instruction in content area classes, to help improve literacy. In this view, subject area teachers should adopt general reading and writing strategies from the language classroom in their own subjects.

However, this approach has proven to be insufficient (O'Brien, et al., 1995). Over the past decades, educational researchers such as Moje (2008) have argued that secondary content literacy should focus on subject areas, instead of on general literacy. They called for a change: content should always be put first, since learning in a subject area entails the understanding of the norms of practice for producing and communicating knowledge in that specific discipline. Disciplines can thus be considered discourse communities which students must navigate. This calls for disciplinary strategies, which should be emanated from content.

In the Netherlands, it is not a shared opinion among teachers that literacy should be part of every subject area, as an intrinsic element of the subject curriculum. Literacy development is still mainly considered to be the responsibility of the language department. History teachers share this latter view, even though it is well-known that history is a subject which demands extensive reading and writing (Mottart et al., 2009).

In the current paper, we report on the design and evaluation process of a lesson series aimed at secondary school (11th grade) students' historical writing, with one focus in particular: learning remains paramount. With this study, we aimed to contribute to educational research by providing insights in effective elements of intervention research, by describing design principles and learning activities explicitly, and thoroughly, following a constructive trend initiated by other researchers in the educational domain (e.g., Schrijvers et al., 2019; Van Ockenburg et al., 2021a).

1.1 Relevance of teaching disciplinary literacy

When literacy is overlooked in the content classroom, disciplinary knowledge of writing remains out of sight, although this knowledge of discipline-specificity is crucial (Carter, 2007). For example, history students need to contextualize a phenomenon in their text, use sources to build an argument, and discuss reliability of sources. It is reasonable to expect such a 'historical lens' to be a history teacher's responsibility, since it is inherent to the domain of history.

However, content area teachers barely spend time supporting students' writing process, nor do they provide writing process instruction (De Oliveira, 2011; Gillespie

et al., 2014; Mottart et al., 2009). This may be logical from the content teacher's point of view, since it is not a formal objective of the Dutch history program to teach students how to communicate history (CvTE, 2022). Teachers tend to align their assignments with what is asked in the final exam: in the Netherlands, that is, answering short open-ended questions. These questions generally focus on assessment of isolated history skills, such as use of sources or argumentation, or on metacognitive concepts, such as explanations or causes. As a result, assignments comparable to the Document Based Question (DBQ), which is an exam task in the US, are much less common.

Yet, also in other countries, teachers struggle with discipline-specific literacy instruction (Nokes, 2010; Ragland, 2007). One explanation might be that teachers feel uncertain about providing reading-writing instruction, since it has had no prominent place in their teacher training either; literacy support seems underestimated and underrated (O'Brien, et al., 1995).

Meanwhile, teachers generally do require their students to write elaborated and well-structured texts. Students who can communicate their thoughts clearly, are more convincing in showing their disciplinary reasoning skills than students who cannot. It is thus an advantage for students to be able to diligently transform thoughts into text, which in itself should already be a solid reason to include disciplinary literacy in the classroom.

Moreover, the act of writing might enhance the learning process. Several studies have suggested that writing can become a powerful means of retrieving, rethinking, revising, and reformulating what one knows (e.g., Galbraith, 1999; Graham et al., 2020; Klein, 1999). In current practice, writing assignments are not always designed as writing-to-learn tasks due to a lack of knowledge about the kinds of learning that writing might engender (Newell & Winograd, 1995). When a specific task is well considered, however, a writing assignment might serve this additional goal of learning, which is the most prominent objective of content classrooms.

1.2 Historical writing

Writing in history is an eminently useful learning activity. Wiley and colleagues (2014) defined learning history as the attempt 'to try to understand the past'. Historical reasoning and problem solving often center on texts. 'Understanding the past' might thus be interpreted as the construction of mental models of past phenomena, based on various documents representing diverse perspectives. This is a learning goal that writing could support perfectly.

Historical writing entails that students must use historical evidence, drawn critically from primary source documents, to write well-structured and well-substantiated arguments (De La Paz & Felton, 2010). It should be distinguished from generic argumentation, as historical arguments require "conceptual understanding, procedural knowledge of historical analysis, an underlying grasp of the topic and discipline, and background content knowledge" (Monte-Sano, 2010, p. 560).

According to Geisler (1994), expertise in a certain discipline requires familiarity with its *content* and its *rhetoric*. For historical writing, novice history students may see historical knowledge as random bits to be reproduced, summarized and written down in a preconceived format to relate what they know about the topic. In contrast, expert historians would see historical knowledge as a construction of an evidenced interpretation of an issue and use the rhetorical strategies of the disciplinary genre to transform contrasting bits of information into a coherent text (McCarthy Young & Leinhardt, 1998). This framework is consistent with Scardamalia and Bereiter's (1987) groundwork on novices and experts in writing (knowledge *telling* versus knowledge *transformation*). In upper secondary education, students will most likely not reach an expert level, especially when it comes to rhetoric, since this aspect is largely underexposed in history education when compared to content (Geisler, 1994).

1.3 Learner characteristics

When designing an instructional unit aimed at historical writing, it should be taken into account that two learner characteristics might have a moderating effect. Firstly, students with high writing *self-efficacy*, generally show low writing apprehension and enjoy writing more, which is associated with higher text quality (Pajares & Valiante, 1997; Sanders-Reio et al., 2014). Secondly, students may hold implicit *writing beliefs* which relate to writing quality (White & Bruning, 2005). These beliefs comprise what students think proficient writing is, and what they think good writers do; these beliefs thus modify what their composing process will look like, and how the eventual text will be established (Graham et al., 1993). White and Bruning (2005) identified *transmissional* and *transactional* beliefs. Writers with high *transmissional* beliefs generally view writing as a means for reporting 'someone else's facts'. These writers stay close to the information and arguments they find in source materials. In contrast, those writers with high *transactional* beliefs, are cognitively and emotionally engaged in their writing processes. They see writing as a means of developing their understanding of the issues at stake, and their own views on these issues. In our instructional unit, we aimed to appeal to students with all self-efficacy levels and all writing belief profiles.

1.4 Research questions and design research methodology

Following McKenney and Reeves' generic model for design in research (2019), we conducted a design study that sought to establish design principles for the integration of disciplinary writing into history, and the development of sound lesson materials for history teachers. For high quality interventions, Nieveen (1999) proposed four generic criteria: content validity, construct validity, practicality, and effectiveness.

Our main research question was therefore:

What is a valid, practical and effective design for writing tasks, and writing instructions, for upper secondary school history to support students' historical writing?

McKenney and Reeves distinguished three phases in design research. During the analysis and exploration phase (1), a literature review and an explorative preliminary context study were conducted. During the design and construction phase (2), the design principles were formulated, and a conceptual model was created. Thirdly, the evaluation and reflection phase (3) involved a trial intervention study.

Our research questions for the different phases were:

- RQ phase 1 How is students' writing in history currently addressed by history teachers, which design requirements can be derived from these findings, and which design principles can be derived for approaching historical literacy? (i.e., focus on validity)
- RQ phase 2 How can these design principles be translated into an instructional unit? (i.e., focus on validity and practicality)
- RQ phase 3 How do students and teachers interact with the instructional unit, in what ways is this different from intended interactions, with what results on historical writing, why does this seem to be the case, and what is the role of learner characteristics (writing beliefs and self-efficacy)? (i.e., focus on practicality and effectiveness)

In this work, we will report on the method and outcomes for each of these three phases sequentially. However, it must be noted that the design process was iterative, and thus the three phases interacted.

2. ANALYSIS AND EXPLORATION PHASE

2.1 Explorations

To explore the context, we interviewed ten history teachers from different schools in the Netherlands (age ranged from 32 to 63 ($M=42$)). These teachers were all qualified teachers of upper secondary levels (years of experience ranged from 4 to 35 ($M=16$)). Each interview took about one hour, and all was audio-taped and subsequently transformed into a written protocol. The aim of the interviews was to obtain a clear view of teachers' current writing and support practices.

We used a stimulated recall strategy: teachers were asked to bring a writing task, containing multiple sources, which they had used recently in an upper secondary class, and two example texts (one weak and one strong) from their students. This task, with accompanying example texts, was used as an angle for the interviews. We discussed five themes: task characteristics, assessment criteria, support practices, cognitive processes, and writing beliefs. For cognitive processes, we asked teachers to explain which processes students should perform, to complete the assignment successfully. Teachers wrote these processes on separate sticky notes, to

subsequently paste and organize them on a piece of paper and draw the connections between the processes. For the full instructions for the interviewees and the full interview guide see supplementary materials. The written protocols were analyzed per theme, with a specific lens for design requirements and principles, using a content analysis procedure (Krippendorff, 2013).

Furthermore, we conducted a literature search. We searched for English journal articles describing writing interventions in history at secondary school levels (e.g., De La Paz, 2005; De La Paz & Felton, 2010; De La Paz et al., 2017; Graham & Perin, 2007; Martinez et al., 2015; Monte-Sano, 2011; Reynolds & Perin, 2009; Van Drie et al., 2006, 2014, 2015). The aim of this search was to discover how historical writing was addressed in previous studies, and with which results for historical writing, to eventually derive *design requirements* and *design principles*, which could aid our instructional design. These principles are generally considered to be the core of the instructional unit and are therefore described to profoundly contribute to theoretical understanding (Rijlaarsdam et al., 2017).

2.2 Interview results

In this section, the interview results which were viewed most relevant for our design are concisely presented. First of all, with regard to support practices, our results confirmed findings in previous studies (e.g., De Oliveira, 2011): teachers rarely mentioned writing support practices, implying that writing support was not often provided, at least not consciously. Four out of ten teachers did not mention writing support practices at all. While the other six teachers did, this was mostly related to the intended text; for example, teachers explained how the eventual text should be structured, or which criteria were considered to be important. Three teachers did mention practices related to the process, for example, discussing how to assess source texts, or modelling how to deduce arguments from source texts. One of these three teachers showed awareness of the relevance of process instruction:

"If there's a source text, you read the question first. You then read the question again, and then you start thinking: what is being asked? And only then do you look in a source for the answer." (Teacher 10)

However, if support practices were mentioned, they focused on aspects of reading and analyzing source texts, and not on writing.

Regarding cognitive processes, we found that teachers rarely addressed any aspects of the writing process. Six teachers mentioned the writing process in terms of "writing" or "writing competence".

One teacher, who had learned about writing discourse in a professional learning community, named the writing part of the process more extensively: for example, they mentioned "dividing into paragraphs" and "critical rereading and fine-tuning" as processes. Three teachers mentioned no steps focused on writing at all; they stuck

strictly to historical skills, such as "analyzing sources" and "drawing conclusions," leaving writing implicit.

Nevertheless, all teachers mentioned writing skills as important for the subject of history. Several teachers noted that what students *meant* often differed from what they actually *wrote*. In such cases, students had generally not mastered the content convincingly. Teacher seven elaborated: "If a text is well written, then you immediately believe that the student understands it".

Moreover, five out of ten teachers mentioned the improvement of students' writing as a goal of the writing task they brought to the interview. Two of them considered "learning to express oneself" the main goal; for the other three teachers, writing development was a subgoal next to content knowledge goals. The remaining five teachers mentioned solely content goals as the learning goals for the writing task, and not writing goals.

2.3 Design requirements

Several design requirements were derived from the exploration of the context. The interview results showed that teachers struggled to unequivocally explain which processes their writing tasks implied for their students. A first requirement for the design, therefore, was to consider the fact that history teachers lack knowledge about writing processes.

A second requirement was to keep practicality in mind. Writing instruction need to be easily applicable for history teachers; interview results showed that teachers do not regularly provide students with writing instruction, or support during the writing process (also: De Oliveira, 2011). As teachers commonly find it difficult to teach reading and writing strategies, the instruction must be made as easily integrable as possible, for them to maximize practicality. Practicality can be divided into three criteria, all related to classroom ecology and a teacher's goal system (Westbroek et al., 2020). A first criterion is *instrumentality*: are procedures available to show how the innovation should be implemented? A second criterion is *congruence*: is the innovation sufficiently congruent with regular practices and important goals, that the teacher connects with? And as a third criterion, *low cost*: do the estimated benefits of implementation outweigh the effort it takes to implement the innovation? Keeping track of these practicality criteria is thus strongly recommended for the instructional design.

A third requirement centers around the finding of previous research that there is an ongoing subject culture, with 'content first' as an important element (e.g., O'Brien et al., 1995). Specifically, half of the interviewed teachers not mentioning "writing development" as a learning goal for a writing task, might be seen as a confirmation of this persistent culture. Although our aim was to eventually improve students' historical *literacy*, an important design requirement thus would be that literacy is developed *in service of content*.

2.4 *Design principles*

A literature search to discover how previous studies addressed students' historical writing was conducted, including a search for intervention studies in secondary grades regarding historical writing combined with learning. De La Paz and colleagues (De La Paz & Felton, 2010; De La Paz et al., 2014; De La Paz et al., 2017) had similar objectives in their respective studies, and they reached encouraging results. Explicit instruction in disciplinary thinking which is embedded in reading-writing tasks, based on sources, supported students' historical argument writing. We have derived two design principles from the aforementioned studies by De La Paz, which are further presented and discussed in the subsequent section.

2.4.1 *Design principle #1: Writing task design*

Similar to De La Paz and colleagues, we aimed students to learn through writing. In their studies, students wrote 'historical inquiries', which they explain as 'working with conflicting primary sources to investigate a central question' (De La Paz et al., 2017, p. 36). Several authors (Monte-Sano & De La Paz, 2012; Newell & Winograd, 1995; Van Drie, et al., 2006; Voss & Wiley 1997) have previously explored the task effects on learning; these studies have underscored that the arguing genre is best suited for writing to learn, since this genre will promote deeper processing of content materials.

Van Drie and colleagues (2006) concluded that *evaluative* prompts (e.g., 'Were the changes in the behaviour of Dutch youths in the 1960s revolutionary?') were most powerful for eliciting historical reasoning, compared to an *explanatory* prompt (e.g., 'How can the changes in the behaviour of Dutch youths in the 1960s be explained?'). A think-aloud study (Holdinga et al., 2021) showed that with such evaluative tasks, the writing process itself is likely to enhance students' thought process, which subsequently led to inclusion of this specific type of argument tasks into our design.

To enhance learning, it is advised to use *source-based* writing tasks. The sourcing skill is considered key in history education (Wineburg, 1991), which makes source-based writing a suitable learning activity (Brante & Strømsø, 2018). This was confirmed by a study of Monte-Sano and De La Paz (2012), which revealed that writing tasks which were stimulating engagement in sourcing, corroboration, and causal analysis, improved students' capability of recognizing and reconciling historical perspectives significantly. Furthermore, as Voss and Wiley (1997) concluded, writing arguments from sources might facilitate understanding of content.

Writing tasks should contain *multiple documents* representing multiple perspectives on the issue at stake, since documents are written from an author's perspective, and no single perspective can be an exhaustive presentation of a historical phenomenon (Britt & Rouet, 2012). Additionally, inclusion of *primary*

documents is recommended, since these stimulate students' attention to source characteristics and trustworthiness and make students more likely to include references to source documents in their own writing (Rouet et al., 1996; Mierwald et al., 2022). After all, constructing arguments from several primary sources, is the essence of history; students learn that history is not just about learning names and dates. It is "an on-going debate about what those facts may mean" (Voss & Wiley, 1997, p. 264).

Furthermore, the writing tasks should not result in extensive texts, in order to avoid the risk of discouraging teachers by an overwhelming paper load (Newell & Winograd, 1995). Hence, it is best to use tasks that can be performed within one lesson, resulting in texts of approximately 200-300 words, which is much shorter than the tasks and texts in other studies. Lastly, frequent practicing is preferred (Graham & Harris, 2017), which is also easier achieved with shorter tasks.

To summarize, our first design principle is therefore:

Design principle #1: If we want students to develop a profound understanding of history through writing, then they should write short evaluative texts, based on multiple primary sources which represent multiple perspectives.

2.4.2 *Design principle #2: Writing process support*

A second key element in our design, considers teachers' *support* of students writing. From our exploration of the context, we concluded writing support was absent, or merely focused on the product. Therefore, discipline specific strategy instructions should be included, to support students *during* writing. This strategy instruction should be discipline-specific since this is preferable, exceeding simply adding a general reading-writing strategy instruction to the disciplinary classroom (Gillis, 2014).

The writing part of the reading-writing process should be emphasized, since this part of the process was most disregarded by the interviewed teachers. Monte-Sano and Allen (2018) drew a similar conclusion, as history teachers from their respective studies tended to focus more on the historical work involved in writing, than on the composition of text. However, a design requirement for the strategy instruction, is that it could be implemented by teachers with no, or only basic knowledge, of writing processes.

Furthermore, it is recommended that the instruction should be flexible, in order to match students' writing preferences (Van Ockenburg et al., 2021b), and adaptable to the task, and task difficulty (Holdinga et al., 2021). Students who have more background knowledge, tend to move faster through the process, and start writing sooner in the process compared to students who have less background knowledge (Holdinga et al., 2021). The latter generally stay longer in the phase of reading and planning. Students' writing routines are likely to differ from task to task, and from topic to topic (Van Steendam et al., 2022). Furthermore, research on writing-to-learn about literary texts, has indicated that adapting writing tasks to students' writing

strategies, increases their learning in the field of literature (Kieft et al., 2008). Therefore, we will offer students a dual-route strategy. These two aspects - *focus on writing* and *flexibility* - of the instruction are distinctive to our design.

To summarize, our second design principle is:

Design principle #2: If we want students to develop a profound understanding of history through writing, it is best to provide them with discipline-specific, dual-route, reading-writing strategy instruction which is easily applicable for teachers.

3. DESIGN AND CONSTRUCTION PHASE

In the second design phase, we developed the design principles into a prototype, considering the aforementioned design requirements. The design-as-constructed is presented, and supported by literature, in the following section.

3.1 *Design-as-constructed*

The first design principle has implied the development of a *task* intervention. As learning activities, students perform short evaluative writing tasks. The second design principle had implied an *instructional* intervention, since we want to support students' performance when executing this type of tasks. We have developed a discipline-specific strategy as the object of the instruction. Previous intervention studies on instructional design provided our framework for development of the strategy, and for the instructional design. We will present these two aspects of the intervention in the next sections.

3.1.1 *Development of the strategy*

To decide which strategy was most suited, we conducted a literature analysis on earlier intervention studies which evaluated reading-writing strategies in history. Several studies have evaluated strategies for different parts of the reading-writing process (e.g., De La Paz, 2007; Martinez et al., 2015; Monte-Sano, 2011; Reynolds & Perin, 2009). De La Paz (De La Paz & Felton, 2010; De La Paz et al., 2017) evaluated an *integrated approach*: reading, writing, and history, were taught together, with separate strategies for each phase. In our design, we combined strategies for reading, writing, and historical reasoning, into one overall *Reading-Thinking-Writing* strategy, based on previous literature on reading processes, writing processes, and disciplinary reading and writing. The strategy entails seven steps, which we will substantiate in the following section. The strategy as presented to students, can be found in the supplementary materials.

Monitoring. According to Britt and Rouet (2012), evaluative questions require students to coordinate a series of somewhat iterative steps. Students constantly need to change roles: from reader-thinker to thinker-writer, to writer-reader. These roles are closely interrelated, necessitating flexible processing of each step (Britt &

Rouet, 2012; Rouet & Britt, 2011). Students need to become aware of those role-switches, which can be established by means of monitoring prompts.

Reading. At the start of the process, students explore the task and sources. According to Britt and Rouet (2012), students should establish what they already know, and what they need to know, in order to write a response to the question. Furthermore, students should have a functional understanding of the genre requested in the prompt. Step 1, *analyzing the task*, is thus about constructing a *task model* (Britt & Rouet, 2012). The task model includes a) the task goal, b) actions to achieve the goal, and c) a set of criteria for reaching the goal. In our study, the task goal is to write an evaluative text that takes a stand on a particular controversy, which is supported by discipline-appropriate evidence, but also deals with perspectives that are contrary to the student's stance. At a certain moment during the phase of building a task model, students begin to turn this task goal into a set of actions, in order to establish an educated stance on the issue. Such actions consist of activities such as reading sufficient source information, identifying potential supporting reasons and evidence, selecting strong reasons with respect to available evidence and audience, identifying other-side positions which must be addressed, and identifying appropriate responses or rebuttals for those counterarguments, if possible. Finally, students should identify criteria for accomplishing task goals. To properly turn the task specifications into goals and action plans, students must understand the nouns ('colonial art', 'western museums') and the action words ('should', 'write an argument').

The next step, Step 2, is to *analyze source materials*. For each source, students should generate a *situation model* (Britt & Rouet, 2012; Rouet & Britt, 2011). This model begins with the assumption that texts are experienced as social entities, not simply as a series of linguistic propositions (Wineburg, 1994). Therefore, it is important to detect several features beyond propositional content, such as the author, the genre, the publication date, the intended audience, and the purpose. Understanding of each source text is necessary, to continue the process.

Thinking. Students gradually shift their role from reader to thinker, and to some extent, to writer. Step 3 is to *connect the source information*: in this phase an *intertext model* is created: readers generate links between sources, which can include rhetorical relations such as corroborating, supporting, or contrasting (Britt & Rouet, 2012). Eventually an *integrated mental model* of the situation or phenomenon is created: an internal representation of the situation or phenomenon described across texts (Britt & Rouet, 2012). The structure of the information in this mental model will depend on the content and the nature of the task. The critical point here, is that in a multiple document situation, the reader of the sources is the author of the integrated mental model. This generally requires the content to be *transformed* and to be re-organized. For Step 4, students are advised to *reread the question* to keep track of the focus of the question. This step might be seen as a monitoring prompt.

At Step 5, students begin to *answer the question*. Previous research has shown that we can distinguish different sorts of writing processes (Torrance et al., 1994). Kieft et al. (2008) recommend that writing instruction is adjusted to individual writing strategy preferences. Therefore, in our strategy two routes are amplified in this phase, both well suited for evaluative history tasks: (1) the route of *writing freely*, where students write down all they can think of in a few writing spurts; and (2) the route of *pre-planning*, where students build their text based on a text scheme and use many short writing spurts (Holdinga et al., 2021).

The student should have a clear picture of the main structure of the text by now. An evaluative question generally requires a dual-perspective argumentative text, complemented by a position based on a consideration of all arguments.

Writing. The steps in the final phase concern *communication*. Students will need to continuously switch roles again, from writer-thinker to reader, and back. We distinguish between *revising* (Step 6) and *editing* (Step 7) during this phase. Flower and Hayes (1981) argued that the goal of revision is substantive change: ‘revision can lead to re-seeing, restructuring, even reconceptualizing the entire discourse’ (p. 16). Revision could take place at word level, sentence level, or paragraph level. Especially, free writers will need to invest in extensive revision in order to transform their draft version into a communicative text. The pre-planners will mainly expand their text, from the text plan they developed. The final step for all types of writers, is to edit for language use, minor errors on word or sentence level, and typos.

3.1.2 The ‘how’: Instructional model and key learning activities

Previous studies on intervention research are quite unified in their recommendations for instructional design, as many studies have used the ‘classic’ model of Self-Regulated Strategy Development (SRSD) (Harris & Graham, 1996) as a base. This model consists of six instructional stages with gradual release of support: (1) develop background knowledge, (2) discuss it, (3) model it, (4) memorize it, (5) support it, and (6) independent performance. Several intervention studies aimed at writing-to-learn history have used SRSD, albeit in slightly adjusted versions (e.g., De La Paz & Felton, 2010; De La Paz et al., 2017; Martinez et al., 2015; Reynolds & Perin, 2009). In our study, we followed the model of De La Paz and Felton (2010), who themselves based their design on the cognitive apprenticeship approach to instruction, as developed by Brown et al. (1989). Five stages provided the framework for instruction: *develop background knowledge, describe it, model it, support it, and independent performance*, which was filled in with learning activities (Table 1).

Table 1. Description of instructional stages and learning activities in the design-as-constructed

<i>Lesson*</i>	<i>Instructional stage</i>	<i>Learning activities</i>	<i>Description</i>
<i>T1</i>		Pretest	Prompt: To what extent have the United Nations been successful in the past 50 years?
<i>0</i>	<i>Develop background knowledge</i>	Task experience	Students perform an evaluative writing task to 'experience' the task. The issue in this task is to what extent the Dutch government should return colonial art stolen from local communities in Indonesia during Dutch colonial rule. Two historical and two recent sources are provided representing multiple, contrasting perspectives.
<i>1</i>		Reflection	The students write down what they thought was easy while performing the task on colonial art, and what was difficult.
	<i>Describe it</i>	Building new knowledge	The teacher presents and explains the 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 a modelling peer, who demonstrates 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.
<i>2</i>	<i>Model it (product)</i>	Assessing peers' texts	Students individually assess three exemplars (on the colonial art-task, lesson 0).
		Generating criteria	In a class discussion, students generate a criteria list.
		Applying new learning	Students apply the criteria; they revise the text they wrote about colonial art (lesson 0), with the criteria list in mind.
<i>3 / T2*</i>	<i>Support it</i>	Scaffolded practice	Prompt: To what extent did the images the United States and Soviet Union had of each other play a role in the Vietnam war?

Key Lessons: Strategy Instruction

			Students are scaffolded by the support of the teacher and the written guide. Students are encouraged to collaborate in the prewriting stage. Students hand in their texts to their teacher electronically. The teacher selects three texts for discussion in the next lesson.
4		Teacher feedback	The teacher provides the class with feedback on the task, as he is used to do with similar history assignments.
5 / T3*	<i>Independent performance</i>	Individual work with the help of a written guide	Prompt: To what extent were American and Vietnamese citizens involved in the Vietnam war? Students write their text with help of the written guide. Students hand in their texts to their teacher electronically. The teacher selects three texts for discussion in the next lesson.
6		Teacher feedback	The teacher provides the class with feedback on the task, as he is used to do with similar history assignments.

*The design-as-constructed was implemented as presented in Group 1 only. In Group 2, the key intervention lessons (Lesson 0-1-2) were provided after T2 performance (see also Figure 2).

Note: Lessons of the instructional unit were interspersed with content lessons (except for Lesson 0-1-2, which were taught consecutively).

We developed a series of learning activities, to be spread over seven 50-minute lessons. A paper workbook was developed for all students, which contained each assignment.

The instructional unit starts with an experience of the task in Lesson 0. In this Lesson, students perform an evaluative task on colonial art. The central issue of this task was: 'To what extent do you think Western museums should return colonial art to the country of origin?' Four sources were provided, all textual (mean length: 189 words). Multiple perspectives were represented, from: (a) a Dutch assistant resident in Indonesia (in 1856); (b) a Dutch artist who accompanied the army on a military expedition to Java, to collect Javanese art (around 1906); (c) the current king of Klungkung, Bali (in 2020); and (d) the director of the National Museum of Indonesia (in 2020).

In Lesson 1, students were prompted to reflect on their writing process. Subsequently, they were presented the Read-Think-Write Strategy by direct instruction, and they watched a video showing a modeling peer, who demonstrated each step (Figure 1). Fragments of modeling were alternated with instructional parts, explaining the strategy. The modeling peer was a volunteering 12th grade student, who was pre-informed about the strategy and instructed to demonstrate all steps

while performing the task about colonial art from Lesson 0. It took two sessions of about 40 minutes of filming, to eventually create a video of 12 min 52 s, in which the modelling peer was visible and audible during 7 min 38 s. While watching the video, students performed a compare-contrast assignment to help them reflect on their writing process.

In Lesson 2, students discussed *exemplars*, which were printed in their workbooks. These texts were collected in an earlier stage and written by 11th grade students in response to the task of Lesson 0. Analyzing and discussing exemplars is a means of imparting teachers' tacit knowledge of criteria (Polanyi, 1973). Discussing exemplars has been shown to reduce differences between student and teacher ratings in previous research (Orsmond et al., 2002). The learning activity thus contributes to task representation.

Lessons 3 and 5 contained writing tasks developed by students' teachers, tailored to the regular curriculum. These tasks were also included in the paper workbook, but students were instructed to write their texts on a computer. Similar to Martínez et al. (2015), we added a written guide ('Cheat sheet', see supplementary materials) as an additional support in the *support it* and *independent performance*-stages (Lessons 3-6). In Lesson 4 and 6, the teachers were considered to provide feedback on students' texts as they were used to do when they would provide feedback on regular history assignments.

Figure 1. Video fragments: an instructional slide (on the left) and a still from the modelling peer illustrating the process (on the right). The original video was in Dutch.



In the subsequent phase of the study, we implemented the design-as-constructed into practice, which is described in the next sections.

4. EVALUATION AND REFLECTION PHASE

In the third phase, we implemented the instructional unit into the practice of two history teachers. Method and results of this trial intervention study are presented in the next sections.

4.1 Research design

For the implementation, we used a switching panels design, with three measurement occasions (T1-T2-T3). An overview of the research design is presented in Figure 2. Teacher A implemented the key intervention lessons (Lesson 0-1-2 from Table 1) between T1 and T2, teacher B between T2 and T3. In between lessons of the instructional unit, lessons focused on knowledge building. Content lessons were also provided to prepare students for the historical issues discussed in the writing tasks.

The three measurements were spread over a period of 8 weeks. For ethical reasons—we did not wish to waste lesson time purely for testing—we integrated our measurements into the regular lessons. In Group 1, therefore, T2 was Lesson 3, and T3 was Lesson 5. In Group 2, T3 was Lesson 3. Lessons 5 and 6 were absent in this Group.

The switching panels design provided us with the opportunity to implement the intervention lessons twice in a row (first in Group 1, then in Group 2), with the groups functioning as each other's control group. Moreover, students were not withheld writing instruction with this design, as is ethically preferable (Cohen et al., 2011).

Figure 2. Overview of the research design

		Panel 1			Panel 2		
Group 1 Teacher A n = 31	Research activity	T1	Intervention	T2	Control	T3	
	Lesson		Lesson 0-1-2	Lesson 3	Lesson 4	Lesson 5	Lesson 6
Group 2 Teacher B n = 30	Research activity	T1	Control	T2	Intervention	T3	
	Lesson				Lesson 0-1-2	Lesson 3	Lesson 4

Note: "Lesson" refers to the lessons described in Table 1.

4.2 Participants

Two qualified teachers at the same school implemented the intervention lessons in their 11th grade (16-17 years old, pre-university track) history classes. Teacher A (a 58-year-old male) had 33 years of experience teaching history, while Teacher B (a 42-year-old female) had 17 years, respectively. Neither teacher had any experience in teaching writing. The class of teacher A consisted of 31 students (20 female); while teacher B taught 30 students (11 female). The students were all actively informed about the research procedure and gave passive consent for participation.

4.3 Measurements

4.3.1 Practicality: Reflective interviews

We aimed to gain insight into the functionality of the intervention, by reflective interviews with both teachers and students. The interview guides are added in the supplementary materials. We conducted reflective interviews of approximately 45 minutes with each teacher. We also conducted two 45-minute group interview sessions with students, one session per group, and each time with two students (Group A: Student A1 and A2; Group B: Student B1 and B2). The students were all female and participated voluntarily. All interviews were conducted online (via Microsoft Teams), and audio-recorded, for which all interviewees actively consented. The audio-recordings were subsequently translated into written protocols.

4.3.2 Effectiveness: Text quality

Writing Tasks. To indicate effectiveness, we used three writing tasks which were co-designed with the participating teachers. The tasks fitted into the regular curriculum and replaced regular workbook assignments. Students developed basic content knowledge on the topic of the tasks, before executing each task. The topic of the term was the Vietnam war. Task 1 focused on the question “To what extent have the United Nations been successful in the past 50 years?”. Seven sources were presented: two cartoons, and five textual sources (mean length: 183 words). Task 2 centered around the question “To what extent did the images the United States and Soviet Union had of each other play a role in the Vietnam war?”. Six sources were presented: one cartoon, and five textual sources (mean length: 129 words). Task 3 queried “To what extent were American and Vietnamese citizens involved in the Vietnam war?”. Five sources were presented, all textual (mean length: 180 words). At all three measurements, students were also allowed to use textbook materials. All tasks contained sources which represented multiple perspectives and were predominantly primary.

Rating instrument. We used text quality as an indicator of effectiveness on three dimensions: (1) quality of content, (2) quality of structure, and (3) holistic quality. We used rubrics for the assessment of content quality and structure. The rubric for content quality was based on the studies of De La Paz and colleagues (2017). They proposed four substantive criteria for students’ historical texts—contextualization, rebuttal, substantiation, and perspective recognition—which we adopted, and complemented with additions from several other intervention studies (Britt & Aglinskas, 2002; Martinez et al., 2015; Van Drie et al., 2015; Voss & Wiley, 1997; Young & Leinhardt, 1998). This resulted in an operationalization of the content criterion into the following four aspects: (a) *understanding of the issue*, (b) *multi-perspectivity*, (c) *elaboration of argumentation*, and (d) *use of sources*.

For quality of structure, we focused on the organizational pattern of the text, a decision that was based on previous studies of Young and Leinhardt (1998) and Reynolds and Perin (2009). De La Paz (2010, 2017) did not include structure as a separate criterion; however, their holistic scoring rubrics highlight rhetorical aspects, also with a focus on overall text structure and coherence. The rubric contained two aspects: (a) global and (b) local text structure. The complete rubric for content and structure, can be found in Appendix A.

For assessing students' texts holistically, we constructed a scale with five example texts, exemplifying different levels of quality. For the selection of these five texts, we went through four steps. As a start, one author (LH) selected 30 texts written by students out of a pre-collected pool of 55 texts about the colonial art task; we ascertained that the sample covered three levels of quality: weak (20%), average (60%), and good texts (20%). Subsequently, a random jury panel of 11 teachers was asked to rank these 30 texts using Comproved, an online tool for comparative judgment (<https://www.comproved.com>). Each of these teachers was asked to make 30 comparisons between randomly composed pairs of texts, and to decide which text was better. For each fifth comparison, we asked for an explanation of the teacher's preference. These elaborations gave us the opportunity to refine the scale with annotations which could explain quality differences.

The third step was to select five texts, which were evenly distributed in the rank order that was constructed by the jury. We selected the texts with z-scores closest to -2, -1, 0, +1, and +2. Additionally, we checked if standard errors of these texts were not exceptionally large. This resulted in the selection of the texts from ranks 3 ($z = +1.88$, $se = 0.59$), 6 ($z = +1.07$, $se = 0.52$), 15 ($z = -0.01$, $se = 0.49$), 24 ($z = +0.98$, $se = 0.55$), and 29 ($z = +1.88$, $se = 0.61$), to function as example texts in the scale. The final step was to annotate the scale on the criteria content and structure.

4.3.3 *Learner characteristics: Writing beliefs and self-efficacy questionnaires*

The learner variables were measured by two questionnaires (Appendix B). The validated writing beliefs questionnaire (Vandermeulen et al., 2020) consisted of 26 statements to rate on a five-point Likert scale. We distinguished two scales: a transmission scale ($\alpha = .68$) (e.g., "I write to clarify what others think") and a transaction scale ($\alpha = .74$) (e.g., "Writing helps me see the complexity of ideas") (White & Bruning, 2005). The two beliefs did not correlate ($r = -.08$, $p = 0.61$).

The self-efficacy questionnaire consisted of 30 statements to rate on a scale from 0 ("I cannot do this at all") to 100 ("I can do this perfectly"). We constructed this self-efficacy questionnaire based on a previous study (Holdinga et al., 2021), covering ten aspects involved in historical writing tasks: reading and analyzing sources; connecting information; selecting information; synthesizing; use of sources; content; coherence; structure; language, and monitoring. These aspects were not intended to be separate constructs; we intended to cover the whole reading-writing process as it is involved in an evaluative task. The Cronbach's alpha of the self-efficacy

questionnaire was .95. The mean score of the 30 items functioned as self-efficacy score.

4.3.4 *Fidelity measures*

To check whether the implementation of the intervention was indeed executed as intended, we used O'Donnell's (2008) five criteria for measuring fidelity of implementation. The first criterion is (a) *adherence*; to check whether the components of the intervention were delivered as designed, the teachers completed teacher logs via email. After each interventional lesson, the teachers updated the team on what was done in the lesson, what plans they had for the next lesson, and what adjustments they had made in the lessons as designed, and in their plans. We also conducted reflective interviews, and one of teacher A's lessons was observed by LH. In the reflective interviews with teachers and students, we also checked for (b) *duration*—the number, length, and frequency of implemented lessons, (c) the *quality of delivery*, (d) *participant responsiveness*, and (e) *program differentiation*—whether critical features that distinguish the program from the comparison condition were present during implementation.

4.4 *Data collection*

4.4.1 *Procedures and circumstances during the intervention period*

Unfortunately, the research was conducted during a period of several (partial) CoVid-lockdowns, and therefore there were restrictions to adapt to. The project started while the country was in a lockdown situation, during which only fully online secondary education was provided, i.e., the teacher taught from home, via Microsoft Teams. After a few weeks, just before Task 2, the situation changed into a hybrid situation, i.e., the teacher taught from school, with half of the students present in the classroom, and the other half of the students attending online. The two groups (online-offline) alternated every full day, so students were physically at school five days out of every two weeks. Groups were split randomly. Group 1 received the strategy instruction in a full online setting; Group 2 in a hybrid situation.

During the regular knowledge building lessons, the teachers explained history content in a traditional way: the teacher explained the course, as well as causes of the Vietnam war. The teachers interacted with the students now and then, in order to keep them alert, but on the whole, the students mainly listened and took notes. This was alternated with short tasks, such as drawing timetables and discussing maps of the Vietnam War situation.

4.4.2 *Procedures and circumstances during measurements*

The teacher supervised the writing tasks and questionnaires, which were performed on a computer, and were handed in digitally, using an electronic learning system. All students performed the questionnaires and Task 1 online, while at home. Tasks 2 and 3 were performed in the hybrid situation, with half of the students at home, and the other half in a computer classroom. Supervision varied in the online situations; students who were at home were not always willing to put their cameras/microphones on. Most texts (80%) were written on the computer, some texts were written by hand (20%), photographed, and handed in digitally. The number of handwritten texts was larger in Group 1 (31% of the total number of texts was written by hand) than in Group 2 (6% handwritten texts). For the analysis, all handwritten texts were transformed into typed texts.

4.4.3 *Rating procedures*

The data set consisted of 121 texts (T1: 49, T2: 36, T3: 36) with three different topics, which were randomly divided per topic, over seven text sets (each containing the three topics; approximately 17 texts per topic). A panel of seven raters evaluated the texts: all raters were qualified current or former teachers of history. Each rater rated three different text sets (total approximately 50 texts), partially overlapping with four other raters, resulting in all texts receiving three rates by three different raters for all three variables. The intraclass correlation coefficients of these three measures were .79 for the holistic score, .73 for content, and .77 for structure. Further analyses were conducted using the means of the three ratings.

4.4.4 *Implementation fidelity*

From the fidelity measures, several deviations from the design-as-constructed were observed. Firstly, from the interviews, we noted that for teacher B, the model texts remained underexposed in the implementation, potentially since they did not feel confident teaching this particular assignment. Students in Group 2 did examine the model texts, but they were then not further discussed with the teacher (as occurred in Group 1). The quality of delivery might thus have been lower in Group 2.

Secondly, feedback lessons in both groups were short (feedback took about 10 minutes of the lesson), and these were not focused on students' writing processes. The teacher manual did not provide strict guidance here; teachers were free to discuss the tasks at their discretion. They were encouraged to refer to the strategy whenever possible, but in practice, teachers only discussed *content*. For example, they highlighted which elements in the text were necessary to discuss the task topic properly.

We also signaled some deviations from the design-as-constructed that can be directly linked to the teaching being in an online setting. For instance, online

education impeded interaction with students. For example, the practice task from Lesson 3 (see Table 1) was intended as a scaffolded task, but because of the online education, students performed the writing task mostly individually, without much scaffolding from peers or their teacher. Although it was possible to consult the teacher, practice showed us that students hardly asked questions in online situations. Furthermore, the individual periods were shortened from 50 to 40 minutes quite ad hoc. This caused teacher A to spread Lesson 1 from the intervention over two periods, to meet the teaching goals.

Moreover, participant responsiveness was only moderate. For example, in Group 1, 55% completed all three tasks, while in Group 2 this was 30%. The merit for students to hand in their work declined over the tasks: in Group 1 the response rate for T1-T2-T3 was respectively 84%, 65%, 70%, and in Group 2 this was 77%, 53%, 47%. A possible explanation for this decline, could be students' demotivation caused by the ongoing online class situation. In a short questionnaire, students were afterwards asked why they had not handed in their work in time. They self-reported several reasons: they said they had 'had no time for it', they had 'forgotten to do it', or they had 'not considered it a priority'. These remarks indicated that many students did not finish the tasks during the lesson time which was appointed to the performance of the task by the teachers. Both teachers noted in their evaluations, that the students' overall motivation declined during the lockdown period, and that it was a challenge to get students motivated in general.

4.5 Data analysis

4.5.1 Analyzing text quality

Indicators for text quality were: (1) holistic quality; (2) quality of content; (3) quality of structure, and (4) text length. Specifically, text length was considered to be an indicator of text quality, as good writers generally write longer texts (Ferrari et al., 1998).

The scores on the criteria content, and on the structure, correlated substantially (T1: $r = .89$; T2: $r = .80$; T3: $r = .83$). Both structure ($r = .94/.89/.92$, $p < .01$) and content ($r = .93/.92/.89$, $p < .01$) also correlated strongly with the holistic score. Text length correlated with holistic quality, at .77, .82 and .87, respectively ($p < .01$). Despite these correlations, we nonetheless report further analyses for all indicators, as due to the small scale of this trial study, all possible influential factors should be examined.

For the analyses of students' scores, we conducted a multilevel analysis. Effects of the interventions in both panels were tested by comparing four nested models: Model 0 with Subject as random factor, Model 1 with the added factor Time, Model 2 with the added factor Condition, and Model 3 with the added interaction Time*Condition. We set the alpha level on $p < .10$, to avoid a Type II error in the case

of a small sample. The comparisons of the fit of the nested models are presented in Appendix C.

4.5.2 Analyzing effect of moderator variables

To explore moderating effects of learner variables (beliefs and self-efficacy), we first centered the scores for three potential moderators, and then extended the models presented above, with Model 4 (general effect of the learner variable LV); Model 5 (LV*Time); Model 6 (LV*Condition), and Model 7 (LV*Time*Condition).

Four students showed an outlier score (outside 95% the confidence interval of the mean) on one of the belief variables (high on Transaction 2x, low on Transaction 1x; low on self-efficacy 1x). These cases were not included in analyses in which belief-variables were explored for moderating a condition effect.

4.6 Preliminary analysis

Prior to the intervention, the two groups did not differ with regard to beliefs and self-efficacy (Pillai's Trace .111, $F(3,40) = .434$, $p = .730$). Students' mean scores are presented in Table 2. The mean score for self-efficacy (70 out of 100) showed that on average, participants found themselves sufficiently capable for the type of tasks we asked them to perform.

To check whether data loss during the intervention was non-systematic in terms of the three belief scores, we ran multivariate analyses for T2 and T3. At T2 no effects of groups were observed (Pillai's Trace .069, $F(3,29) = .714$, $p = .552$). At T3, however, a multivariate effect was observed (Pillai's Trace .271, $F(3,27) = 3.353$, $p = .034$). Subsequent univariate analyses showed a significant difference between groups for transactional beliefs at Task 3; Group 1 scored significantly higher than Group 2 (3.4 vs 2.8, $F(1,29) = 9.499$, $p = .004$).

Table 2. Mean beliefs and self-efficacy scores

	Group 1 N = 23		Group 2 N = 21	
	M	SD	M	SD
Transmissional beliefs	3.19	.52	3.15	.43
Transactional beliefs	3.29	.46	3.11	.65
Self-efficacy	70.15	13.02	69.00	10.40

5. RESULTS

In this section, results of the trial intervention are presented on three criteria of high-quality interventions: validity, practicality and effectiveness.

5.1 Validity and practicality

5.1.1 Teachers' perspective

Overall, teachers were positive about the tasks; they would surely reuse them in their teaching, also in other grades, because they felt that such tasks opened doors for a discussion of relevant historical issues, they wanted their students to become aware of:

Teacher A: *"For me, this is the essence of history, that you can have very different views on facts and events, based on factual material, and interpretations."*

The teachers were less convinced they would reuse the strategy instruction, although they were both highly positive about the usefulness of the strategy itself, and they would re-use that in their lessons.

The video evoked divergent reactions. The modeling peer, and the explanation of the strategy, were useful elements, but they suggested that the duration of the explanatory part of the video could be shorter, in order to improve ease-of-use for students. Moreover, as the video was built upon the task regarding colonial art, they felt this to hinder transferability to other grades or topics. Teacher A was most positive about the video:

Teacher A: *"I noticed the students were very interested to see how someone who is struggling, I think that's worth a million, to demonstrate those steps so visually."*

Noticeably, both teachers seemed very selective in their enthusiasm about the possible reuse of intervention materials. Especially the aspects about the writing processes (Steps 5, 6, and 7 of the strategy) seemed to lack relevance in their opinion. They wondered whether the steps on writing might be shortened, or even be omitted completely.

Teacher A: "As a historian, I think: I have finished step 1-5 of that strategy, and that's enough, but there are another two steps on that sheet for some reason. Irritating."

Teacher B: "About the writing part, I thought: wow, that's pretty extensive. Could that be shortened? I believe so. I'm not sure, though."

5.1.2 Students' perspective

Students perceived the writing tasks as useful and valuable, and they would like to do such tasks more often:

Student A2: "I liked the tasks. With workbook questions, it's often just a matter of reading comprehension. Questions are asked, and then it's just literally in the text. But now, you are really learning to think."

Although the tasks had increased the students' self-perceived knowledge, students did experience them as "a lot of work". With shorter tasks, with for example fewer

source texts, students indicated they would feel less obstructed, or discouraged, by the estimated workload.

In the students' opinion, the strategy was very extensive; they thought the assignment would take too long, if they would perform all the steps. Some steps were "obvious": students felt they already knew them, and already did what was taught. Students liked to compare their own performance to the strategy, but they were not too excited about the video. The strategy was also explained on paper, so the explanatory part of the video felt redundant, and the modeling peer was too studious. To conclude: they thought of the video as 'dull' and 'slow'.

The model texts were appreciated; these provided the students with a good representation of what an evaluative text might look like, and the students said the model texts were helpful to become aware of the necessary components:

Student A2: "Through model texts, you learn which things are important in a text, and which things you should not do, which you otherwise might accidentally do yourself."

5.2 Effectiveness

For *holistic quality*, *structure*, and *text length*, we did not find interaction effects of Time and Condition, as we had expected. Model 3, which indicates such an interaction effect, fitted only the scores for *content*. Content score results are presented in Figure 3. However, upon exploring differences between groups at each measurement occasion, we found no significant results (T1: mean difference = .153, $se = .244$, $p = .532$; T2 mean difference = .342, $se = .300$, $p = .262$; T3 mean difference = -.442, $se = .297$, $p = .146$).

To point out the scale of the differences between the groups at each measurement occasion, Figure 4 shows the effect sizes. At T3, the level of difference between the groups was .51, which might be considered a 'moderate' effect size.

Figure 3. Content score results (Scale 1-5)

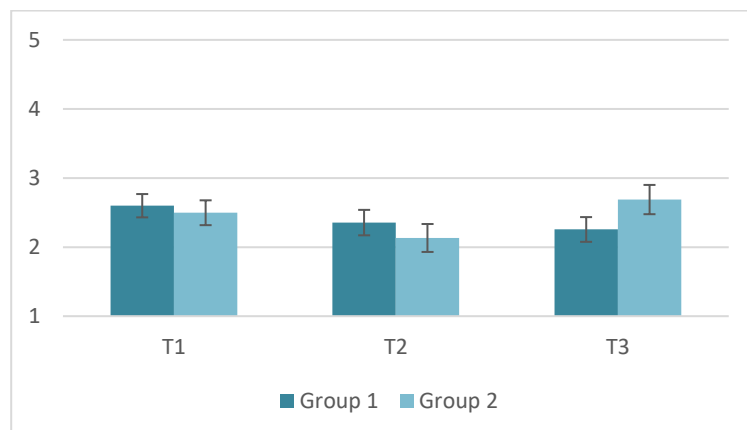
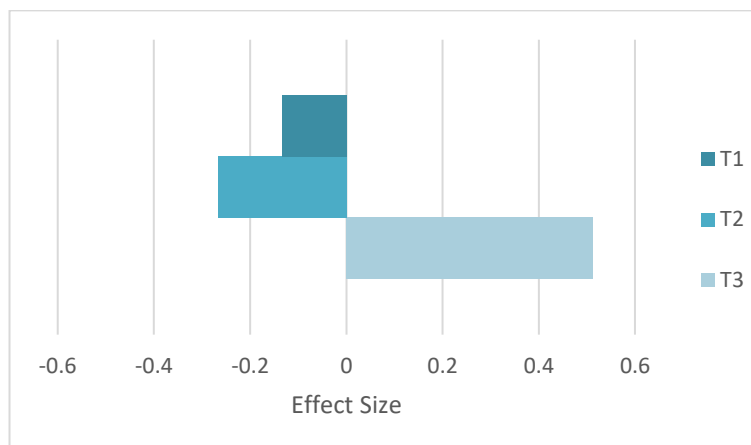


Figure 4. Effect sizes for content scores (Reference = Group 1)



5.3 Moderating effects of learner variables

We tested if one of the learner variables moderated the intervention effect. No moderation was revealed, for none of the three learner variables, and for none of the dependent variables.

What we did observe, with one exception¹, was that adding learner variables in Model 4 resulted in a better fit (Appendix C) for self-efficacy, and for transactional beliefs, indicating direct effects of these factors. In all cases, the effect of self-efficacy was positive, varying from $\beta = .03$ ($p = .02$) for content, $.03$ ($p = .03$) for structure, and $.48$ ($p = .03$) for holistic quality. The effect of transactional beliefs was negative (content: $-.82$, $p = .004$; structure: -0.75 , $p = .02$; text length -109 , $p = .002$; holistic quality: -15 , $p = .01$).

5.4 Effect of online/offline

In the second panel, a new variable was introduced: the hybrid situation, during which students were assigned into off- or online participation, by the school. For none of the four dependent variables, an effect of this instruction condition was observed.

6. DISCUSSION

While every discipline has its own crucial specificities when it comes to reading and writing in upper secondary grades, the overarching question remains how best to support students in acquiring these skills. In the current study, we have described

¹ The exception was self-efficacy in the analysis for text length.

our instructional design process, to develop materials that would help teachers improve 11th grade students' historical writing. In this paper, we have reported on the design process, which contained three phases: (1) an analysis and exploration phase, (2) a design and construction phase, and (3) an evaluation and reflection phase.

In the first phase, we have explored the educational context through an interview study, resulting in three design requirements: it is advised to take into account (a) teachers' lack of knowledge about writing processes, (b) practicality demands, and (c) the subject culture of 'content first'. Furthermore, a literature search led to two design principles for our instructional design:

Design principle #1: If we want students to develop a profound understanding of history through writing, then they should write short evaluative texts, based on multiple primary sources that represent multiple perspectives.

Design principle #2: If we want students to develop a profound understanding of history through writing, it is best to provide them with discipline-specific, dual-route, reading-writing strategy instruction which is easily applicable for teachers.

In the second design phase, together with two history teachers each teaching an 11th-grade class, we co-developed three evaluative writing tasks, to be used as a replacement of regular workbook exercises. Additionally, we developed a strategy on how to perform evaluative writing tasks, which was based on previous studies on reading and writing processes. Our aim was that students would be able to oversee the historical issue at stake, by means of the evaluative prompt, *and* be able to translate these insights into text. The design process resulted in tangible lesson materials for 11th grade students, in order to improve students' historical writing.

In the third phase, we implemented the constructed design in two groups, and we evaluated its validity, practicality, and effectiveness. We explored effectiveness of the tasks, and the instructions, on students' historical writing, for which we assessed students' texts holistically, on content, structure, and text length.

In the implementation of the design, Principle 1 stood out clearly; reflective interview results indicated that the evaluative writing tasks were perceived as useful learning activities by both students and teachers. Principle 2, the strategy instruction, stood out less clearly. Aspects of the instruction manual were underrepresented, or underrated. The strategy elements focusing on *communicating* knowledge, which were so valued from a literacy development perspective, were not fully internalized by the teachers. We consider this our main concern for redesign: to challenge teachers' beliefs system.

With regard to effectiveness, analyses of students' text quality indicated an effect of strategy instruction on *content quality*, however, only in Group 2. With the use of evaluative writing tasks combined with strategy instruction, Group 2 students' content knowledge was better presented in their texts. We found no effects on holistic scores, structure, and text length.

Learner variables might have influenced the effects on quality. We have conducted the intervention during a challenging educational setting. Although we

did not find relations between students' beliefs and quality directly, we did find moderating effects of transactional beliefs. In general, students were demotivated to participate in online class settings. Among students of Group 2 with higher transactional beliefs scores, the motivation was deemed even lower, while we expected these students to be the more persistent ones, as they were conceiving the writing tasks as useful for learning, and thus scoring higher than average.

One influential factor on learning outcomes might have been motivation. As Pajares (2003) concluded, students' self-efficacy in writing can influence writing motivation as well as writing outcomes. Our data seemed to confirm this, since we found a direct positive effect of self-efficacy on text quality. However, we did not find a moderating effect of self-efficacy.

Another influential factor might have been the amount of writing tasks in the relatively short intervention period of eight weeks. With four writing tasks in two months, students might have felt overloaded. As one student stated in the evaluative interviews: "We were not used to such demanding tasks in history. The tasks took quite a lot of effort" (Student B2).

Next to that, it remains unclear to what extent the combination of the different writing tasks influenced students' historical writing development. From the instructional unit, in which an example task on colonial art was used, we assumed students to transfer their newly obtained reading-writing strategy knowledge to the teacher-designed writing tasks, which were embedded in content lessons. It is possible, however, that all writing tasks, including the pretest, had influence on students' disciplinary writing. Furthermore, students' initial writing proficiency might have been of influence as well.

To conclude, learning outcomes are promising, yet diffuse. This might be problematic: when a unit is proven effective, this might convince teachers to implement the materials. However, even though the learning outcomes of this study are encouraging, they might not convince teachers completely yet. Therefore, recommendations for redesign are presented in the next section.

6.1 Recommendations for redesign

Overall, the functionality of the learning materials seemed satisfactory. Both students and teachers were particularly positive about the type of writing tasks. Therefore, recommendations for redesign mainly concern the strategy instruction.

First, from the interviews with students and teachers, we conclude the video needs improvement in pace and variety. The explanatory parts can be shortened, and the video could show more than one student, for example. Also, we recommend stretching the intervention over a larger time frame, to avoid students feeling overloaded with extensive writing, which might cause motivational problems.

A next recommendation is more fundamental, considering teachers' awareness of design principles and supporting theories, as well as teachers' knowledge. Teachers, in general, are often faced with unexpected events in the classroom,

forcing them to adapt their lesson plan to the situation. Teachers must then be able to decide which learning activity can be adapted, and in what way, to still meet the lesson objectives. If the teacher does not have a good understanding of the reasoning behind a certain learning activity, there is a risk of conducting, or adapting, the learning activity in a non-constructive way: it may become much less meaningful. Previous research has shown that individuals rely on heuristics, when situations become complex, and time and information are limited (Westbroek et al., 2020). Thus, a teacher with a clear understanding of his own heuristics, and of the goals behind the material should be the goal. This might be achieved by sharing design principles explicitly. In our study, we mainly informed the teachers about the practical aspects of intervention, and not about our design principles. The teachers were involved in the development of the writing tasks, but not in the development of the strategy instruction. It was remarkable that teachers considered the possibility of shortening, or even deleting, the strategy steps on writing (Steps 5-6-7), since we actually saw these as crucial elements of the intervention. This sense of necessity obviously did not come across, and it clearly showed the knowledge gap we had already reported on in the exploration of the context. Professional development of teachers in these fields thus seems necessary and has proven its use in previous studies (e.g., Van Drie et al., 2017).

A final recommendation for redesign, is to supply more guidance on how to provide feedback on students' texts. In our study, we left the feedback options open, which led to a focus on content only. When feedback is directed more towards the writing process itself, this might reinforce the strategy instruction even more.

6.2 *Limitations and strengths*

From our evaluations, we consider two elements of the materials and design process, to be particularly successful: (1) the writing tasks and (2) the co-development of the tasks. The writing tasks used in this study were evaluative in nature and based on sources. These tasks were considered instructive and reflecting "the essence of history". Writing thus fitted well with the goals of the subject of history. Furthermore, the tasks were constructed by the teachers themselves. This ensured teacher involvement, and a good fit with the curriculum, which are factors that increase the likelihood of reuse.

Although common for a design study, a next limitation of the trial study was its relatively small scale, and that it was conducted at only one school. Other limitations regard the circumstances; due to the online educational setting, teachers' supervision while students performed the tasks was not optimal. It resulted, for example, in students writing by hand instead of using the computer. This was not preferred, since previous meta-analyses (Bangert-Drowns, 1993; Goldberg et al., 2003) concluded that texts composed on a computer are generally of greater length, and of higher compositional quality, than handwritten texts. Furthermore, a major disadvantage was that students were not able to write *in* the classroom, in teacher's

sight. This eliminated the general idea behind the inclusion of writing tasks; as supporting students' processes while writing was almost not possible, and students could not collaborate during the pre-writing stage. We thus recommend the trial study to be replicated under more conventional circumstances, with regard to educational practice in the future. In such a follow-up study, we might want to include measurements for content learning, and the influence of initial writing performance, to obtain more insights in learning outcomes.

Despite the small scale of the study, we did find effects on possibly the most important quality measure: content. In the rubric, content quality contained four factors: understanding of the issue, multiperspectivity, argumentation, and use of sources. These factors are important in learning history, and therefore improvements in these factors are obviously imperative for history teachers. Considering the subject's culture of 'content first', this result might be a prerequisite for teachers to actually work on literacy.

Moreover, this encouraging outcome was the result of only a short intervention, consisting of merely three lessons. This was a conscious choice; we aimed to design lessons which were feasible and had a low threshold for practical implementations. It seems promising that effects can already be achieved with relatively little time investment for history teachers; after all, practicality was one of the design requirements. A potential follow-up study could be to design a viable and effective professionalization course for teachers, which will not heighten the threshold for implementation of writing instruction into teachers' subject area.

The most important strength of our study might be the set of validated design principles, and intervention materials, for teachers and future educational researchers with a focus on disciplinary literacy. The design principles may offer guidance on how to integrate discipline-specific writing instruction into subject areas. The description of the intervention of Table 1, is a tangible aid for history teachers worldwide, on how writing instruction might be implemented into the history classroom, in only a few lessons. The strategy we developed was based on previous studies on historical reading and writing research and is usable for students in upper secondary levels. We have attempted to describe the theories behind the principles and learning materials as thoroughly as possible, and we hope to inspire teachers around the world to integrate discipline-specific writing instruction into their teaching.

Apart from these practical implications, this design study also has the potential of strengthening overall knowledge on historical literacy. We have tried to develop a complete, yet compact, discipline-specific reading-thinking-writing strategy, building upon previous interventional research. We may contribute to the educational research field with insights into teachers' views on their own role in developing students' historical writing, and on the role of literacy in the history classroom in general.

ACKNOWLEDGEMENTS

We would like to thank the students and teachers who participated in this study greatly for their time and effort.

DECLARATION OF CONFLICTING INTERESTS

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

FUNDING

This research was partly funded by the Dudoc-Alfa Sustainable Humanities Programme (project DA2-2018-14).

REFERENCES

- Bangert-Drowns, R. L. (1993). The word processor as an instructional tool: A meta-analysis of word processing in writing instruction. *Review of Educational Research*, 63(1), 69–93. <https://doi.org/10.3102/00346543063001069>.
- Brante, E. W., & Strømsø, H. I. (2018). Sourcing in text comprehension: A review of interventions targeting sourcing skills. *Educational Psychology Review*, 30(3), 773–799. <https://doi.org/10.1007/s10648-017-9421-7>
- Britt, M. A., & Aglinskias, C. (2002). Improving students' ability to identify and use source information. *Cognition and instruction*, 20(4), 485–522. https://doi.org/10.1207/S1532690XCI2004_2
- Britt, M. A., & Rouet, J. F. (2012). Learning with multiple documents: Component skills and their acquisition. In J. R. Kirby, & M. J. Lawson (Eds.), *Enhancing the quality of learning: Dispositions, instruction, and learning processes* (pp. 276–314). Cambridge University Press.
- Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, 18(1), 32–42. <https://doi.org/10.3102/0013189X018001032>
- Carter, M. (2007). Ways of knowing, doing, and writing in the disciplines. *College Composition and Communication*, 385–418. <https://www.jstor.org/stable/20456952>
- Cohen, L., Manion, L., & Morrison, K. (2011). *Research methods in education*. Seventh edition. London: Routledge.
- CvTE, College voor Toetsing en Examens [Dutch Institute for Assessment and Exams] (2020). Syllabus centraal examen 2022. *Geschiedenis VWO [Syllabus central exam 2022. History pre-university track]*. Utrecht: College voor Toetsing en Examens.
- 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, 139–156. <https://doi.org/10.1037/0022-0663.97.2.139>
- De La Paz, S., & Felton, M. K. (2010). Reading and writing from multiple source documents in history: Effects of strategy instruction with low to average high school writers. *Contemporary Educational Psychology*, 35, 174–192. <https://doi.org/10.1016/j.cedpsych.2010.03.001>
- De La Paz, S., Felton, M., Monte-Sano, C., Croninger, R., Jackson, C., Shim Deogracias, J., & Polk Hoffman, B. (2014). Developing historical reading and writing with adolescent readers: Effects on student learning. *Theory and Research in Social Education*, 42(2), 228–274. <https://doi.org/10.1080/00933104.2014.908754>
- De La Paz, S., Monte-Sano, C., Felton, M., Croninger, R., Jackson, C. & Worland Piantedosi, K. (2017). A historical writing apprenticeship for adolescents: Integrating disciplinary learning with cognitive strategies. *Reading Research Quarterly*, 52(1), 31–52. <https://doi.org/10.1002/rrq.147>

- De Oliveira, L. C. (2011). Knowing and writing school history: The language of students' expository writing and teachers' expectations. Charlotte, NC: IAP.
- Ferrari, M., Bouffard, T., & Rainville, L. (1998). What makes a good writer? Differences in good and poor writers' self-regulation of writing. *Instructional Science*, 26(6), 473-488. <https://doi.org/10.1023/A:1003202412203>
- Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication*, 32(4), 365-387. <https://doi.org/10.2307/356600>
- Galbraith, D. (1999). Writing as a knowledge constituting process. In G. Rijlaarsdam, E. Espéret, M. Torrance, & D. Galbraith (Eds.), *Studies in writing: Vol. 4. Knowing what to write: Conceptual processes in text production* (pp. 139-160). Amsterdam: Amsterdam University Press.
- Geisler, C. (1994). Academic literacy and the nature of expertise: Reading, writing, and knowing in academic philosophy. Hillsdale, NJ: Lawrence Erlbaum. <https://doi.org/10.4324/9780203812174>
- Gillespie, A., Graham, S., Kihara, S., & Hebert, M. (2014). High school teachers' use of writing to support students' learning: A national survey. *Reading & Writing*, 27(6), 1043-1072. <https://doi.org/10.1007/s11145-013-9494-8>
- Gillis, V. (2014). Disciplinary literacy: Adapt not adopt. *Journal of Adolescent & Adult Literacy*, 57(8), 514-623. <https://doi.org/10.1002/JAAL.301>
- Goldberg, A., Russell, M., Cook, A., & Russell, E. M. (2003). The effect of computers on student writing: A meta-analysis of studies from 1992 to 2002. *Journal of Technology, Learning, and Assessment*, 2(1), 1-52. <https://ejournals.bc.edu/index.php/jtla/article/view/1661>
- Graham, S., & Harris, K. R. (2017). Evidence-based writing practices: A meta-analysis of existing meta-analyses. In R. Fidalgo & T. Olive (Series Eds.) & R. Fidalgo, K. R. Harris, & M. Braaksma (Vol. Eds.), *Studies in Writing Series: Vol. 34. Design Principles for Teaching Effective Writing* (pp. 13-37). Leiden: Brill.
- Graham, S., Kihara, S. A., & MacKay, M. (2020). The effects of writing on learning in science, social studies, and mathematics: A meta-analysis. *Review of Educational Research*, 90(2), 179-226. <https://doi.org/10.3102/0034654320914744>
- Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology*, 99(3), 445-476.
- Graham, S., Schwartz, S. S., & MacArthur, C. A. (1993). Knowledge on writing and the composing process, attitude toward writing, and self-efficacy for students with and without learning disabilities. *Journal of Learning Disabilities*, 26(4), 237-249. <https://doi.org/10.1177/002221949302600404>
- Harris, K., & Graham, S. (1996). *Making the writing process work: Strategies for composition and self-regulation* (2nd ed.). Cambridge: Brookline Books.
- Holdinga, C. C., Janssen, T.M., & Rijlaarsdam, G. C. W. (2021). The relationship between students' writing process, text quality, and thought process quality in 11th-grade history and philosophy assignments. *Written Communication*, 38(4) 544-586. <https://doi.org/10.1177/07410883211028853>
- Kieft, M., Rijlaarsdam, G., & Van den Bergh, H. (2008). An aptitude-treatment interaction approach to writing-to-learn. *Learning and Instruction*, 18(4), 379-390. <https://doi.org/10.1016/j.learninstruc.2007.07.004>
- Klein, P. D. (1999). Reopening inquiry into cognitive processes in writing-to-learn. *Educational Psychology Review*, 11(3), 203-270. <https://doi.org/10.1023/A:1021913217147>
- Krippendorff, K. (2013). *Content analysis. An introduction to its methodology*. Third Edition. LA: Sage Publications.
- Martínez, I., Mateos, M. D. M., Martín, E., & Rijlaarsdam, G. (2015). Learning history by composing synthesis texts: Effects of an instructional programme on learning, reading and writing processes, and text quality. *Journal of Writing Research*, 7(2), 275-302. <https://doi.org/10.17239/jowr-2015-07-02-03>
- McCarthy Young, K. M., & Leinhardt, G. (1998). Writing from primary documents: A way of knowing in history. *Written Communication*, 15(1), 25-68. <https://doi.org/10.1177/0741088398015001002>
- McKenney, S., & Reeves, T. C. (2019). *Conducting educational design research*. Second edition. Oxon/New York: Routledge.

- Mierwald, M., Lehmann, T., & Brauch, N. (2022). Writing about the past: The impact of different authentic instructional material on students' argument writing in history. *European Journal of Psychology of Education, 37*, 163–184. <https://doi.org/10.1007/s10212-021-00541-5>
- Moje, E. B. (2008). Foregrounding the disciplines in secondary teaching and learning: A call for change. *Journal of Adolescent & Adult Literacy, 52*, 97-107. <https://doi.org/10.1598/JAAL.52.2.1>
- Monte-Sano, C. (2010). Disciplinary literacy in history: An exploration of the historical nature of adolescents' writing. *Journal of Learning Sciences, 19*(4), 539-568. <https://doi.org/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. <https://doi.org/10.1111/j.1467-873X.2011.00547.x>
- Monte-Sano, C., & Allen, A. (2018). Historical argument writing: the role of interpretive work, argument type, and classroom instruction. *Reading & Writing, 32*(6), 1383-1410. <https://doi.org/10.1007/s11145-018-9891-0>
- Monte-Sano, C., & De La Paz, S. (2012). Using writing tasks to elicit adolescents' historical reasoning. *Journal of Literacy Research, 44*(3) 273-299. <https://doi.org/10.1177/1086296X12450445>
- Mottart, A., Vanbrabant, P., & Van de Ven, P.-H. (2009). Schrijven bij diverse schoolvakken. Een verkenning. [Writing in different subjects. An exploration.]. *Levende Talen Tijdschrift, 10*(4), 14-21.
- Newell, G. E. & Winograd, P. (1995). Writing about and learning from history texts: The effects of task and academic ability. *Research in the Teaching of English, 19*(2), 133-163. <https://www.jstor.org/stable/40171246>
- Nieveen, N. (1999). Prototyping to reach product quality. In J. van den Akker, R.M. Branch, K. Gustafson, N. Nieveen, & T. Plomp (Eds.), *Design approaches and tools in education and training* (pp. 125-136). Boston: Kluwer Academic.
- Nokes, J. D. (2010). Observing literacy practices in history classrooms. *Theory & Research in Social Education, 38*(4), 515–544. <https://doi.org/10.1080/00933104.2010.10473438>
- O'Brien, D. G., Stewart, R. A., & Moje, E. B. (1995). Why content literacy is difficult to infuse into the secondary school: Complexities of curriculum, pedagogy, and school culture. *Reading Research Quarterly, 30*(3), 442-463. <https://doi.org/10.2307/747625>
- O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K–12 curriculum intervention research. *Review of Educational Research, 78*(1), 33-84. <https://doi.org/10.3102/0034654307313793>
- Orsmond, P., Merry, S., & Reiling, K. (2002). The use of exemplars and student derived marking criteria in peer and self-assessment. *Assessment & Evaluation in Higher Education, 27*, 309–23.
- Pajares, F. (2003). Self-efficacy beliefs, motivation, and achievement in writing: A review of the literature. *Reading & Writing Quarterly, 19*(2), 139-158. <https://doi.org/10.1080/10573560308222>
- Pajares, F., & Valiante, G. (1997). Influence of self-efficacy on elementary students' writing. *The Journal of Educational Research, 90*(6), 353-360. <https://doi.org/10.1080/00220671.1997.10544593>
- Polanyi, M. (1973). *Personal knowledge*. London: Routledge/Kegan Paul.
- Ragland, R.G. (2007). Changing secondary teachers' views of teaching American history. *The History Teacher, 40*(2), 219–246.
- Reynolds, G. A., & Perin, D. (2009). A comparison of text structure and self-regulated writing strategies for composing from sources by middle school students. *Reading Psychology, 30*, 265-300. <https://doi.org/10.1080/02702710802411547>
- Rijlaarsdam G., Janssen, T., Rietdijk, S., & Van Weijen, D. (2017). Reporting design principles for effective instruction of writing: Interventions as constructs. In R. Fidalgo & T. Olive (Series Eds.) & R. Fidalgo, K. R. Harris, & M. Braaksma (Vol. Eds.), *Studies in Writing Series: Vol. 34. Design Principles for Teaching Effective Writing* (pp. 280-313). Leiden: Brill.
- 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. <https://doi.org/10.1037/0022-0663.88.3.478>
- Rouet, J. F., & Britt, M. A. (2011). Relevance processes in multiple document comprehension. In M. T. McCrudden, J. P. Magliano, & G. Schraw (Eds.), *Text Relevance and Learning from Text* (pp. 19-52). Greenwich, CT: Information Age Publishing.

- Sanders-Reio, J., Alexander, P. A., Reio Jr., T. G., & Newman, I. (2014). Do students' beliefs about writing relate to their writing self-efficacy, apprehension, and performance? *Learning and Instruction*, 33, 1-11. <https://doi.org/10.1016/j.learninstruc.2014.02.001>
- Scardamalia, M., & Bereiter, C. (1987). Knowledge telling and knowledge transforming. *Advances in Applied Psycholinguistics*, 2(2), 142-175.
- Schrijvers, M., Janssen, T., Fialho, O., & Rijlaarsdam, G. (2019). Toward the validation of a literature classroom intervention to foster adolescents' insight into human nature: an iterative design process. *L1-Educational Studies in Language and Literature*, 19(4), 1-47. <https://doi.org/10.17239/L1ESLL-2019.19.04.02>
- Torrance, M., Thomas, G. V., & Robinson, E. J. (1994). The writing strategies of graduate research students in the social sciences. *Higher Education*, 27, 379-392. <https://doi.org/10.1007/BF03179901>
- Van Drie, J., Braaksma, M., & Van Boxtel, C. (2015). Writing in history: Effects of writing instruction on historical reasoning and text quality. *Journal of Writing Research*, 7(1), 123-156. <https://doi.org/10.17239/jowr-2015.07.01.06>
- Van Drie, J., Janssen, T., & Groenendijk, T. (2017). Improving writing in social studies through professional development: Effects on teachers' beliefs, classroom practice and students' writing. *European Journal of Applied Linguistics*, 5(2), 273-308. <https://doi.org/10.1515/eujal-2017-0009>
- Van Drie, J., Van Boxtel, C., & Braaksma, M. (2014). Writing to engage students in historical reasoning. In G. Rijlaarsdam (Series Ed.), & P. D. Klein, P. Boscolo, L. C. Kirkpatrick, & C. Gelati (Vol. Eds.), *Studies in writing: Vol. 28, Writing as a learning activity* (pp. 94-119). Leiden: Brill.
- 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, & G. Erkens, (Eds.), *Collaborative learning, reasoning and technology* (pp. 265-296). Mahwah, NJ: Lawrence Erlbaum Associates.
- Van Ockenburg, L., Van Weijen, D., & Rijlaarsdam, G. (2021a). Learning how to synthesize: The design and evaluation of a reading-writing learning unit for high-school students. *L1-Educational Studies in Language and Literature*, 21(1), 1-33. <https://doi.org/10.17239/L1ESLL-2021.21.01.06>
- Van Ockenburg, L., Van Weijen, D., & Rijlaarsdam, G. (2021b). Choosing how to plan informative synthesis texts: Effects of strategy-based interventions on overall text quality. *Reading & Writing*, 36, 997-1023. <https://doi.org/10.1007/s11145-021-10226-6>
- Van Steendam, E., Vandermeulen, N., De Maeyer, S., Lesterhuis, M., Van den Bergh, H., & Rijlaarsdam, G. (2022). How students perform synthesis tasks: an empirical study into dynamic process configurations. *Journal of Educational Psychology*, 114(8), 1773-1800. <https://doi.org/10.1037/edu0000755>
- Vandermeulen, N., De Maeyer, S., Van Steendam, E., Lesterhuis, M., Van den Bergh, H., & Rijlaarsdam, G. C. W. (2020). Mapping synthesis writing in various levels of Dutch upper-secondary education: A national baseline study on text quality, writing process and students' perspectives on writing. *Pedagogische Studiën*, 97(2), p. 187-236.
- Voss, J. F., & Wiley, J. (1997). Developing understanding while writing essays in history. *International Journal of Educational Research*, 27, 255-265. [https://doi.org/10.1016/S0883-0355\(97\)89733-9](https://doi.org/10.1016/S0883-0355(97)89733-9)
- Westbroek, H., Janssen, F., Mathijssen, I., & Doyle, W. (2020). Teachers as researchers and the issue of practicality. *European Journal of Teacher Education*, 1-17. <https://doi.org/10.1080/02619768.2020.1803268>
- White, M. J., & Bruning, R. (2005). Implicit writing beliefs and their relation to writing quality. *Contemporary Educational Psychology*, 30(2), 166-189. <https://doi.org/10.1016/j.cedpsych.2004.07.002>
- Wiley, J., Steffens, B., Britt, M. A., & Griffin, T. D. (2014). Writing to learn from multiple-source inquiry activities in history. In G. Rijlaarsdam (Series Ed.) & P.D. Klein, P. Boscolo, L. C. Kirkpatrick, & C. Gelati (Vol. Eds), *Studies in Writing: Vol. 28, Writing as a learning activity* (pp. 120-148). Leiden: Brill.
- 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(1), 73-87. <https://doi.org/10.1037/0022-0663.83.1.73>
- Wineburg, S. S. (1994). The cognitive representation of historical texts. In J. Castellan, D. B. Pisoni, & G. Potts (Eds.), *Teaching and Learning in History* (pp. 85-135). Hillsdale, NJ: Lawrence Erlbaum Associates.

APPENDICES

*Appendix A: Rubrics for assessment of content and structure**Table A1. Rubric for assessment of content quality*

Instruction: Give one score from 1-5 for the aspect 'content'.

	1	2	Score 3	4	5
Understanding of the issue	The text barely displays understanding of the main issue, or understanding is lacking completely		The text displays a simple and basic understanding of the main issue		The text displays a thorough understanding of the main issue. This understanding exceeds the specific information in the documents; in the text, own knowledge and information from the source materials are intertwined.
Multiperspectivity	The text is one-sided; the main issue is discussed, but not from different perspectives.		Opposing views are presented in the text, but not very elaboratively		Opposing views are presented in the text
Argumentation	The arguments in the text are weak and/or barely elaborated on		The different views are supported by strong and/or weaker arguments, and the arguments are only moderately elaborated on		The different views are supported by strong arguments, which are elaborated on with accurate facts, examples and details
Use of sources	Source information is presented as 'own text' (e.g., it is presented as facts, and/or references are not provided)		Source information is made explicit in the text by referencing (e.g., 'According to Eduard Douwes Dekker ...')		Source information is integrated in the text; authors of sources are not only mentioned, but also discussed (i.e., the author's position as a reporter)

Table A2. Rubric for assessment of structure

Instruction: Give one score from 1-5 for the aspect 'structure'.

	1	Score 3	5
Global text structure	Overall, the text has no clear structure, which makes the text difficult to understand	Overall, the text has a reasonably clear and logical structure. The text is divided into paragraphs, but those are not always logical or are not presented in the right order.	Overall, the text has a clear and logical structure. The text has a strong introduction and conclusion and is divided into clear paragraphs.
Local text structure	The text is incoherent. The text presents paragraphs or sentences in an unclear manner, because structural elements (signal words, linking words) are lacking. Because of this, the text seems to be a series of sentences without coherence.	Paragraphs are reasonably unified. Linkages between paragraphs or between sentences are mostly implicit, e.g., because linking words or signal words are lacking.	The paragraphs are unified and coherent, both internally and from paragraph to paragraph. Transitions are clear, i.e., through the use of connectors.

*Appendix B: Writing beliefs and self-efficacy questionnaires**Table B1. Writing beliefs questionnaire: Subscale and items*

Transmissional beliefs (9 items, $\alpha = .68$)

1. Good writers include a lot of quotes from authorities in their writing;
 2. The most important goal of writing is to convey information clearly to the readers;
 4. Writing should focus on the information in books and articles;
 5. The key to successful writing is accurately reporting what authorities think;
 6. The most important reason to write is to report what authorities think about a subject;
 20. It is important to objectively convey source information;
 21. It is important to accurately convey source information;
 22. It is important to write objectively;
 23. I write to clarify what others think.
-

Transactional beliefs (8 items, $\alpha = .74$)

8. Writing is a process involving a lot of emotion;
 12. Writing is often an intense experience;
 13. Writing helps me to understand better what I'm thinking about;
 15. Writing helps me see the complexity of ideas;
 16. My thoughts and ideas become more clear to me as I write and rewrite;
 18. Writing is often an emotional experience;
 25. I get lost in writing;
 26. Writing asks the utmost of me.
-

Table B2. Self-efficacy questionnaire (30 items, $\alpha = .95$)

1. I can derive from the assignment what the main historical issue is;
 2. I can purposefully read a source;
 3. I can analyze sources for reliability and representativeness;
 4. I can organize source information;
 5. I can determine whether sources contradict each other or whether they agree;
 6. I can identify different perspectives in the source information;
 7. I can explain different viewpoints by when and where each text was written;
 8. I can identify information that is relevant for the issue;
-

-
9. I can select information that is relevant for my text;
 10. I can write a text that corresponds with the source information;
 11. I can make connections between information from different sources;
 12. I can synthesize information from different sources into one sentence or paragraph;
 13. I can explain source information in my own words;
 14. I can write a text that is comprehensible for someone who has not read the source materials;
 15. I can describe an historical event as an introduction for my text;
 16. I can use historical concepts in my writing;
 17. I can write arguments supported by evidence;
 18. I can explain and illustrate source information in my own text;
 19. I can write a text with a clear main idea;
 20. I can write a text where all paragraphs are useful;
 21. I can write a text without redundant sentences;
 22. I can use conjunctions to link sentences and paragraphs;
 23. I can write a well-structured text;
 24. I can write a text with coherent paragraphs;
 25. I can vary my sentence structures and choice of words;
 26. I can write a text without errors and typos;
 27. I can identify and correct language errors;
 28. I can divide my time between reading and writing well;
 29. I can monitor my process while performing a reading-writing task;
 30. During the process, I sometimes decide to go back to an earlier stage in the process.
-

*Appendix C: Model comparisons for the four quality variables*Holistic quality*Table C1. Multilevel analysis holistic quality*

<i>Models</i>	χ^2	<i>N</i> <i>parameters</i>	<i>Comparison</i>		
			<i>Models</i>	χ^2	<i>df</i> <i>p</i>
Model 0	Intercept + 2 variance components	3			
Model 1	Plus: Time	5	0 vs 1		2 .627
Model 2	Plus: Condition	6	1 vs 2	0.016	1 .901
Model 3	Plus: Interaction Time*Condition	8	2 vs 3	4.062	2 .133

Table C2. Mean holistic quality (and standard errors), estimated under model 3:

<i>Condition</i>	<i>T1</i>	<i>T2</i>	<i>T3</i>
Group 1 (E-C)	84.1 (3.6)	81.8 (3.9)	79.4 (3.8)
Group 2 (C-E)	82.5 (3.8)	79.0 (4.3)	87.3 (4.5)

Content

Table C3. Multilevel analysis content quality

Models	χ^2	N parameters	Comparison		
			Models	χ^2	df p
Model 0		3			
Model 1	282.673	5	0 vs 1		2 .110
Model 2	282.672	6	1 vs 2	0.001	1 .977
Model 3	276.644	8	2 vs 3	6.028	2 .050

Table C4. Mean content quality (and standard errors), estimated under model 3:

	T1	T2	T3
Group 1 (E-C)	2.60 (0.17)	2.36 (0.18)	2.26 (0.18)
Group 2 (C-E)	2.50 (0.18)	2.13 (0.20)	2.69 (0.21)

Structure

Table C5. Multilevel analysis text structure quality

Models	χ^2	N parameters	Comparison		
			Models	χ^2	df p
Model 0		3			
Model 1	323.145	5	0 vs 1		2 .047
Model 2	322.214	6	1 vs 2	0.931	1 .336
Model 3	320.308	8	2 vs 3	1.906	2 .387

Table C6. Mean text structure quality (and standard errors), estimated under model 3:

Condition	T1	T2	T3
Group 1 (E-C)	2.74 (0.20)	2.29 (0.22)	2.44 (0.21)
Group 2 (C-E)	2.35 (0.21)	2.01 (0.24)	2.48 (0.25)

Text length

Table C7. Multilevel analysis text length

Models	χ^2	N parameters	Comparison			
			Models	χ^2	df	p
Model 0		3				
Model 1	1466.586	5	0 vs 1		2	.019
Model 2	1466.571	6	1 vs 2	0.001	1	.903
Model 3	1464.652	8	2 vs 3	6.028	2	.384

Table C8. Mean text length (and standard errors), estimated under model 3:

Condition	T1	T2	T3
Group 1 (E-C)	234 (22.8)	170 (24.9)	175 (24.2)
Group 2 (C-E)	207 (24.3)	186 (27.4)	189 (28.7)

SUPPLEMENTARY MATERIALS

How to answer a 'to what extent' question

When you try to answer a 'to what extent' question, there are several steps to follow. While going through these steps, you should continually monitor what you are doing. While you're reading, ask yourself: Do I need to keep reading or can I start writing? And while writing, ask: Do I need to go back to the source text, or do I know enough to keep writing? In other words, the three steps partially overlap.



READING

Step 1: Read and analyze the question

Determine:

What is the main issue?

When was this issue relevant?

What historical event is most important?

What is the historical context?

Underline key words.

Ask yourself:

What do I already know about this subject?

Step 2: Read and analyze the sources

First skim the source texts, and for each text: identify whose opinion it represents (Which group(s) of people held this point of view? How widely held was it?)

Determine how reliable it is (Can you use this source to prove your point? Who wrote it? Who was the target audience? What did the author want the text to achieve? When was it written? Where did the author's information come from?)

Then read each source fully and carefully. Underline any information that is relevant to the question.



THINKING

Step 3: Make connections

Organize the information from the different sources. What different points of view can you identify? What information goes with what viewpoint? Do the source texts contradict each other or do they agree?

Can the differences between viewpoints be explained by when and where each text was written?

Step 4: Reread the question

Step 5: Answer the question

Route A: Just start writing. Your first draft does not have to be perfect; you can rearrange and add to your text later.

Route B: Plan your text. Structure your information. To answer a 'to what extent' question, you need to discuss at least two different viewpoints.



WRITING

Step 6: Expand your text

Route A: Revise your draft version. Rearrange, structure and add more information.

Route B: Write your text according to your plan.

Pay attention to the following aspects:

Introduce the issue at the beginning of your text. Key words come in handy here. Those are the words you underlined in Step 1;

Signpost: use conjunctions to link sentences and paragraphs. The most common conjunctions used when writing about history indicate change and continuity, similarities and differences, and cause and effect.

Use historical concepts in your writing

Step 7: Check and revise

Reread the question again and then reread your answer. Imagine you are the reader. Are your reasoning and language clear?

Pay attention to detail. Reread your text again and correct any errors and typos.



MONITORING

Throughout the process, keep asking yourself: can I move on to the next step? Or do I need to loop back to the previous step?

Instructions and Interview Guide for the Preliminary Study

Instructions for Interviewees

Please bring the following three items to the interview:

- A short writing task that you have used recently in an upper secondary class. Make sure it meets the following criteria:
 - The task resulted in a text of at least one paragraph (ca. 80 words) or more;
 - The task is source-based, preferably based on multiple sources;
 - You think the task is relevant for your subject area;
 - The task is generally considered challenging by students;
 - The students have performed the task during school hours (not as homework).
- Two example texts from your students: bring one weak text and one good text;
- When available: your assessment rubric for this task.

Interview Guide

- 1) Analysis of the writing task
 - What is the reason you chose to bring this specific task to this interview? In what way is this task relevant for your subject area?;
 - Who developed this task?;
 - Could this task also be a part of an assessment?;
 - What are the characteristics of the source texts? (primary/secondary, genre, length, difficulty);
 - What are the characteristics of the text students wrote? (length, genre, audience);
 - What is the learning goal of the task?;
 - To what extent have students achieved this learning goal?;
 - Were you satisfied with students' results?
- 2) Analysis of students' texts
 - What is it that makes the weak text so weak?;
 - Although you judge the weak text as weak, is there anything good in it?;
 - What is it that makes the good text so good?;
 - What could be improved in the good text?;
 - To what extent is this task discipline-specific? What is typical history in this task?
- 3) Support and feedback
 - What did you and the students do prior to task performance? (instruction, support);
 - What did you and the students do during the task performance? (collaboration with peers, help from the teacher or from tools, time spent, questions asked);

- What did you and the students do after task performance? (grading or not, oral or written feedback from teacher or peers, assessment rubric available or not);
 - In retrospect: what would you do differently prior, during or after task performance?;
 - What might help the student who wrote the good text improve? (prior, during, after performance);
 - What might help the student who wrote the weak text improve? (prior, during, after performance)
- 4) Cognitive processes
- What cognitive processes are involved in the task? Please write them on sticky notes;
 - Can you describe how a student should perform the task, subsequently, from beginning to end? Please, paste the sticky notes in the right order or way to represent the process as a whole, as a flow chart;
 - Could the job be done any other way, different from how you have described the process until now?;
 - What is the main thing you would like to see your students do differently in their process?;
 - How could you take care of that?
- 5) Writing beliefs
- How important is it for your students to have good writing skills?

Interview Guide for Reflective Interviews of the Trial Study

Reflective Interviews with Teachers

- 1) Descriptive
- What did you do in the regular lessons? Please, provide a description of the learning activities during these lessons;
 - How did you conduct the intervention lessons? Did you skip or add elements? Why?
- 2) Evaluative
- To what extent do you think the lesson materials are practical? What improvements do you see?;
 - To what extent do you think the lesson materials are useful for learning? What improvements do you see?;
 - Which elements would you reuse?;
 - How did students respond to the lessons? Do you think the learning goal of writing better historical texts was achieved?

Reflective Interviews with Students

- 1) Program differentiation

- Were the history lessons any different than you were used to?
What was different? Was that positive or negative?
- 2) Motivation
- How motivated were you during the intervention lessons? Why?;
 - How motivated were your classmates?
- 3) Evaluative
- How did you experience performing evaluative tasks?;
 - What did you think of the intervention lessons?;
 - What did you learn from the intervention lessons?;
 - What would you have wanted to learn in addition?;
 - What did you think of the strategy?;
 - Would you keep using the strategy in future assignments?
 - Which step in the strategy was most useful for you?;
 - Did you use the cheat sheet while making the assignments?;
 - What did you think of the video?;
 - What did you think of the modelling peer in the video?;
 - What did you think of the model texts?;
 - What would you tell next years' students in 11th grade who are starting the intervention lessons?;
 - What improvements for the intervention lessons would you suggest?