

Fostering philosophy teachers' disciplinary writing practice: A multiple-case design study

Lieke Holdinga, Jannet van Drie & Gert Rijlaarsdam

University of Amsterdam | The Netherlands

Abstract: In this design study, we designed an instructional unit open to contextual modifications with the aim of fostering secondary school students' philosophical writing. Three philosophy teachers developed innovative source-based writing tasks and provided discipline-specific writing strategy instruction in their 10th grade class.

In this study, we focused on change. We explored teachers' interaction with the instructional design and studied teachers' views on how the intervention had changed their practice since a change of beliefs is crucial to successful, durable innovation of teaching. Moreover, we studied the effects of the changed practice, by exploring change in students' writing. An external jury analyzed students' texts to determine students' actual learning achievements. Teachers' insights into student progress were obtained from reflective interviews that featured comparisons between the observed and expected results.

The results showed that teachers judged the design to be feasible, valid, and effective for students' philosophical writing development. After the intervention, students' texts showed similar or even more independent philosophical thinking than before, while the tasks became more complex. Implementation drove teachers to contemplate writing instruction, indicating a change in their belief system, which is necessary for genuine improvement in teacher practice.

Keywords: disciplinary literacy, philosophy, writing instruction, instructional design, teacher development



Holdinga, L., Van Drie, J., & Rijlaarsdam, G. (2024 - accepted for publication). Fostering philosophy teachers' disciplinary writing practice: A multiple-case design study. **Journal of Writing Research**, volume(issue), #-#-. DOI: xx

Contact: Lieke Holdinga, Research Institute of Child Development and Education, University of Amsterdam, Nieuwe Achtergracht 127, 1018 WS Amsterdam | The Netherlands – c.c.holdinga@uva.nl - ORCID 0000-0002-2543-6594

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

1. Introduction

According to recent international reports, Dutch secondary school students' literacy skills have exhibited an alarming trend of decline (OECD, 2018). Similar trends have been reported in other countries. In the United States, for example, recent national assessment reports have shown that the reading performance of US high school students has not improved since 1971, with only 38% of high school students scoring at or above the level of proficiency (Goldman, 2012). This impoverishment is problematic: literacy skills are a prerequisite for independent participation in our highly literate society, and crucial for students' subject matter learning (Shanahan et al., 2011) and cognitive development (Graham et al., 2020).

The common view of literacy development in recent educational research has been that it is the responsibility of every teacher, not merely that of teachers in the language department. Furthermore, developing literacy is not merely a matter of developing proficiency in general reading and writing practices regardless of subject content (Moje, 2008; Shanahan & Shanahan, 2011). Instead, literacy is now viewed as a crucial element of the task of supporting students in developing knowledge of various subjects. Therefore, literacy instruction should be treated "as a key part of the broader effort" (Heller & Greenleaf, 2007, p. 1).

In the subject of philosophy, which is the context of this study, students are asked to perform a variety of linguistic tasks, which demand a proficient literacy level. Students are required to read and interpret complex primary philosophical texts with the aim of enabling them to answer philosophical questions, which are to be presented in written text or in presentations. Therefore, it might seem obvious for a philosophy teacher to address students' literacy skill development; however, such a focus is not always the case in educational practice. In a previous study on teaching philosophical literacy (Koek, 2020), 90% of philosophy teachers responded that developing their students' literacy was a challenge. With this study, we aimed to support teachers facing this challenge by developing, implementing, and evaluating an instructional unit intended to enhance students' philosophical literacy.

1.1 Teaching Disciplinary Literacy

The integration of literacy into classroom content places a heavy burden on teachers. Teachers who are accustomed to teaching a subject are required not only to have knowledge of the relevant content (i.e., facts, concepts, structural relations, reasoning, and argumentation) but also to obtain knowledge of the reading and writing procedures specific to their discipline (Monte-Sano et al., 2014). Research conducted by Goldman and colleagues (2016) produced a framework featuring five core concepts on which disciplines differ: (1) epistemology; (2) ways of inquiry and

reasoning; (3) concepts, principles, and frameworks for describing and analyzing phenomena; (4) types of sources; and (5) text genres and language use. These concepts are clusters of the types of knowledge that enable subject teachers to formulate learning goals that are targeted at what students need to know and to be able to do attain high levels of literacy in their disciplines. This task might be quite a challenge for teachers who have not been educated to provide instruction on these aspects of literacy.

Teaching materials aimed at the development of students' philosophical reading and writing with the goal of supporting teachers' practice are scarce. For other subjects, enhancing disciplinary literacy has been researched more extensively. For example, for the subject of history by De La Paz and colleagues (e.g., De La Paz & Felton, 2010; De La Paz et al., 2017; Monte-Sano & De La Paz, 2012) and in the fields of math and science by Hand's research team (e.g., Chen et al., 2016; McDermott & Hand, 2016; Villanueva & Hand, 2011). In response to this gap, we developed an instructional unit, with the aim of enhancing students' philosophical literacy. A distinctive aspect of our study was our focus on discipline-specific aspects of philosophical writing processes and texts.

1.2 Enhancing Classroom Practice

To promote proficient student writing, qualified writing tasks and instruction are necessary. This requires teachers to develop knowledge about writing processes and the teaching of writing as well as to reflect on their beliefs concerning the role of writing and the characteristics of disciplinary texts. Teachers who experience effective professional development obtain new knowledge, which changes their beliefs and classroom practice, ultimately fostering student learning (Desimone, 2009).

Previous studies (Samuelowicz & Bain, 1992) have distinguished among five qualitatively different conceptions of teaching, from learning-oriented on the high-quality side of the scale to content-oriented on the low-quality side. Teachers' conceptions of writing are highly correlated with their teaching approaches (Kember, 1997). Ho and colleagues (2001) have recommended a developmental approach that focuses on conceptual change; they argued that real improvement in teaching must begin with a change in the way they think about teaching or, in this case, about teaching writing. Therefore, we designed a teacher guidance program to stimulate contemplation of writing tasks, instruction, and support. This made our research relevant both in a technical way, as teachers changed their practice, and in a cultural way, as teachers challenged their beliefs (De Vries, 1984). We consider these cultural changes to be important with regard to their long-term impact on teacher practice.

1.3 Research Questions

The primary goal of this study was to equip teachers to integrate qualified disciplinary writing instruction into their practice. We conducted a design study that consisted of three phases: (a) analysis and exploration; (b) design and construction; and (c) evaluation and reflection (McKenney & Reeves, 2019). In phases a-b, we focused on preliminary research questions aimed at both designing an instructional unit that can enhance students' philosophical literacy and providing accompanying teacher guidance to ensure the success of the integration and adoption of the unit:

RQ1: Which design principles can be derived from previous research on the development of secondary school students' philosophical literacy?

RQ2: How can these design principles be developed into an instructional unit?

In the third phase, the instructional design was implemented, evaluated, and reflected on by three philosophy teachers, whom we supported by guiding sessions. We explored teachers' interaction with the materials with the goal of understanding how they employed and examined learning activities drawn from the innovative instructional design. The purpose of this exploration was to determine the extent to which teachers' instructional practices had actually changed and the ways in which teachers and students perceived these changes. RQ3 was therefore:

RQ3: How do philosophy teachers interact with (a) principle-based writing task design and (b) an instructional unit intended to support students' disciplinary writing?

Next, we aimed to obtain insights into the perceived effectiveness of this approach with regard to students' writing proficiency and teachers' conceptions of student progress in philosophical writing. RQ4 was therefore:

RQ4: To what extent does the intervention contribute to philosophy teachers' conceptions of progress in students' philosophical writing development?

With regard to RQ4, we asked teachers to define the expected level of student performance, i.e., the performance level that they thought their students were likely to achieve according to their expectations, and we then confronted teachers with the actual level of students (Rijlaarsdam & Janssen, 1996) based on independent assessments of students' writing.

Given the nature of RQ3-4, in the attempt to explore a change in teachers' practices and beliefs and the effects of these changes, we considered multiple-case research featuring predominantly qualitative measures to be an appropriate research approach.

2. Analysis and Exploration Phase

2.1 Exploration of the Context

In a study conducted prior to the intervention, we explored the educational context to ensure that the intervention was suitable for (a) the regular philosophy curriculum, (b) teachers' views on the level of philosophical writing required from students, and (c) teachers' needs for guidance with regard to providing writing instruction. Therefore, we conducted interviews with eleven philosophy teachers drawn from different schools. Interviews took approximately one hour each, were audiotaped, and were subsequently transcribed. We asked teachers to bring an example of a writing task that they had recently used at the upper-secondary level as well as two exemplars, including one weak exemplar and one strong exemplar. Five topics were discussed in the interview: writing task character, assessment criteria, support practices, cognitive processes, and writing beliefs (Appendix A).

The results showed that teachers aimed to provide tasks which stimulated philosophical thinking. Teachers were accustomed to providing feedback on texts and regularly asked students to write several drafts. They varied in terms of their satisfaction with their students' writing proficiency; however, not all students reached the required level. The provision of process instruction or process feedback was scarce. Teachers seemed to lack awareness of writing processes, as they struggled to make explicit the cognitive activities that they assumed students would perform. In conclusion, the starting situation was that while teachers were accustomed to employing writing assignments in their philosophy teaching, they lacked the knowledge and tools necessary to instruct students or support them in the task of writing.

2.2 Design Principles

In a previous study, we established design principles for literacy development in history (Holdinga et al., 2023). Based on the prestudy (§2.1) and a literature search, we adapted these design principles for the instructional unit to the discipline of philosophy:

If we want students to develop a profound understanding of philosophy through writing, then it is best:

1. to use writing tasks that can prompt a discussion of a concise philosophical issue and to accompany this prompt with various (primary) sources that represent multiple perspectives on the issue at hand; and
2. to provide students with discipline-specific, dual-route, reading-writing strategy instruction that is easily applicable for teachers.

We opted to focus on writing-to-learn tasks, since writing has proven its potential for student learning (Graham et al., 2020). Moreover, a combination of learning to write with learning would imply a focus on subject content, which is considered to be an effective feature in educational innovation (Van Veen et al., 2012). In the Netherlands, the main aim of philosophy education is to teach students how to philosophize in their own right. Essay writing can contribute to this goal; it encourages a writer to "explore" and progress through an entire "train of thought" (Velema & Groza, 2020). Philosophical essays are common tasks that teachers use as a learning activity (Marsman, 2010), thus enhancing the feasibility of our design. Therefore, it seems to be best to design tasks that feature a philosophical issue as a prompt. Since reading and writing may reinforce each other (Graham et al., 2018), we recommend the use of tasks based on sources. According to previous research on source-based writing, primary sources representing multiple views on the issue evoke more 'sourcing' and 'referencing' (Britt & Rouet, 2012).

Furthermore, we aimed to emphasize concise tasks, resulting in short texts, for two reasons: (1) we aimed to ensure that students wrote within sight of the teacher to facilitate the provision of process support; and (2) we aimed to enable students to master a procedure and to provide them with multiple opportunities to practice this procedure. Short tasks would then be most efficient. This aspect was a departure from teachers' usual practice: our prestudy showed that philosophical essay tasks in upper secondary grades usually result in long texts (800-1000 words).

We derived the second principle from research on general writing instruction. Providing direct explicit instruction and supporting students' writing processes are known to be effective methods for writing development in general (Graham & Harris, 2017). In another study that focused on the field of history (Holdinga, 2023), we concluded that strategy instruction is also effective for disciplinary writing. In addition, research (Kieft et al., 2008) has shown that adapting writing instruction to students' writing strategies is effective for writing-to-learn, since knowledge construction is best served when students can use their preferred strategy. This insight led us to employ a dual-route strategy.

Previous research has shown that innovative methods should be easily applicable for teachers. Demands regarding practicality (instrumentality, congruence, and low cost; Westbroek et al., 2020) should thus be considered. Accordingly, the unit should be open to contextual modifications. When the design is flexible regarding adaptation to different topics and levels of performance, it drastically enhances the feasibility of this approach for teachers.

Finally, since our prestudy showed that philosophy teachers did not have much experience in teaching writing, we require guiding activities to support the integration and implementation of the unit into teachers' practice. To guide the design, we considered critical features of effective professional development (Desimone, 2009; Van Veen et al., 2012): content focus, active learning, alignment with teachers' goals, and sufficient duration. A final design principle thus focused on teacher guidance:

If we want teachers to design qualified writing tasks and provide writing strategy instruction, then it is best

3. to support teachers' integration and implementation process with guiding activities that can prompt reflection on the part of teachers while considering critical features of effective professional development.

3. Design and Construction Phase

3.1 Writing Task Design

For the benefit of ecological validity and teacher engagement, we entrusted task design to the participating teachers. Therefore, the first principle was operationalized into three requirements: the writing tasks were required to (1) prompt an exploration of a concise philosophical issue, (2) be based on various (primary) sources and represent multiple perspectives, and (3) be appropriate for the duration of one class.

3.2 Writing Strategy Instruction

To elaborate the second principle, we designed (a) a strategy for students that explains how to perform the writing tasks most effectively and (b) an instructional design aimed at teaching students the strategy effectively.

3.2.1 The RTW Strategy

We designed a genre- and discipline-specific Read-Think-Write strategy (RTW strategy), based on studies regarding philosophical reading and writing (Borren, 2012; Concepción, 2004; Corcelles Seuba & Castelló, 2015), multiple text reading (Britt & Rouet, 2012; Rouet & Britt, 2011), knowledge constitution processes (Galbraith, 2009; Galbraith & Baaijen, 2018; Van den Broek & Helder, 2017), and writing processes (Flower & Hayes, 1981; Elbow, 1973).

The RTW strategy is divided roughly into three main cognitive processes: reading, thinking, and writing. Furthermore, students are prompted to monitor the entire process to develop a recursive process. The strategy provided two routes, since previous research has shown that we can recognize and distinguish between different types of writing processes (Torrance et al., 1996). Accordingly, Kieft et al.

(2008) recommended that writing instruction be tailored to individual writing strategy preferences. Consequently, two routes were included in the RTW strategy, to accommodate free writers (route A) and pre-planners (route B).

We aimed to address all five core concepts in which disciplines differ according to the framework developed by Goldman et al. (2016). For example, students were prompted to connect abstract concepts with concrete examples (addressing overarching philosophical frameworks) and to find their own voice in their texts (addressing philosophical epistemology/discourse language). Figure 1 shows the strategy as presented to students.

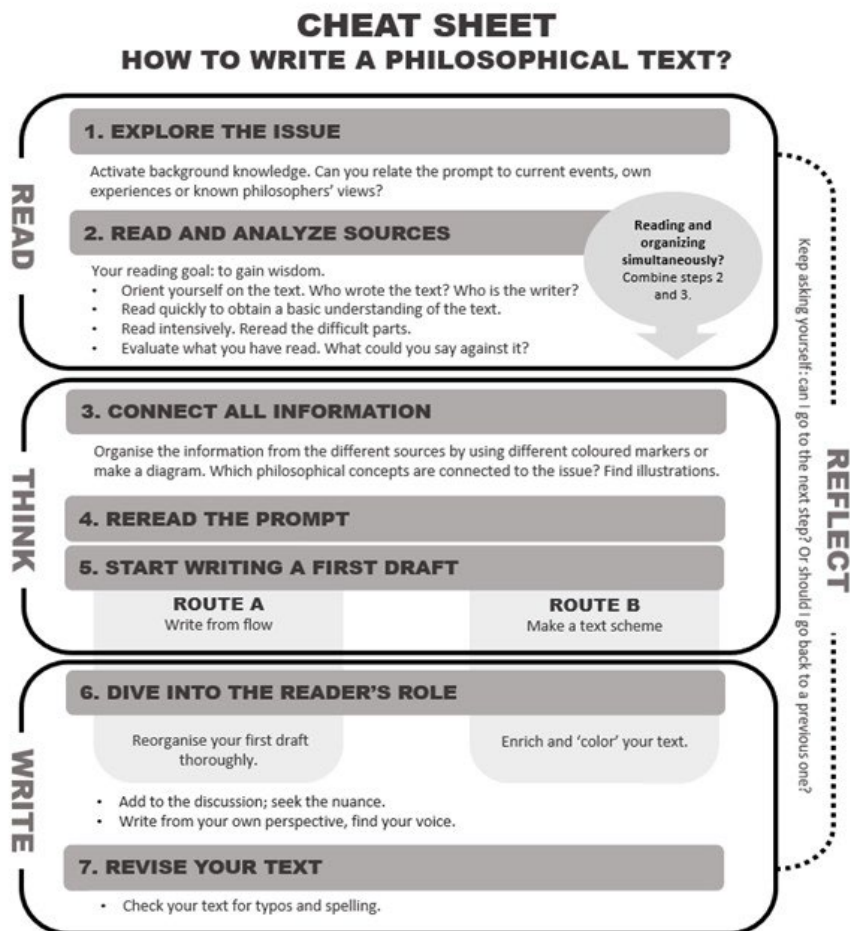


Figure 1: The Strategy as Presented to Students (Concise Translation; the Original "Cheat Sheet" Was More Extended and Was Presented on Folded A5-Wrapper in Full Color)

3.2.2 Instructional Design

To teach students the strategy, we followed the instructional model proposed De La Paz and Felton (2010) to develop students' writing proficiency in the field of history. This model was inspired by the classic model of Self-Regulated Strategy Development (SRSD) presented by Harris and Graham (1996). The framework for instruction contains five stages: develop background knowledge, describe it, model it, support it, and independent performance. We adopted these stages and provided them with content in terms of learning activities. In the current study, the intervention consisted of six lessons (Table 1).

A teaching manual was developed to convey the instructional unit to the teachers. For the intervention lessons, students received a paper workbook containing a first writing task (T1), Lesson 1, and Lesson 2. All three writing tasks were to be completed using a computer.

The instructional unit started with T1 for students to "experience" the task, which activated students' prior knowledge. In Lesson 1, students reflected on this writing experience. Subsequently, students were presented the RTW strategy, which was modeled by peers (12th grade students from other schools) in a video as a mode of observational learning (Braaksma et al., 2004). The video (Figure 2) contained scenes in which students struggle and engage in monitoring. For example, a modeling student concluded that they did not understand a source and started rereading an excerpt. Students performed a compare-contrast assignment while watching the video to help them reflect on their own process. In the teaching manual, teachers were instructed to pause the video after each stage (read-think-write) and to reflect on the models in a class discussion.

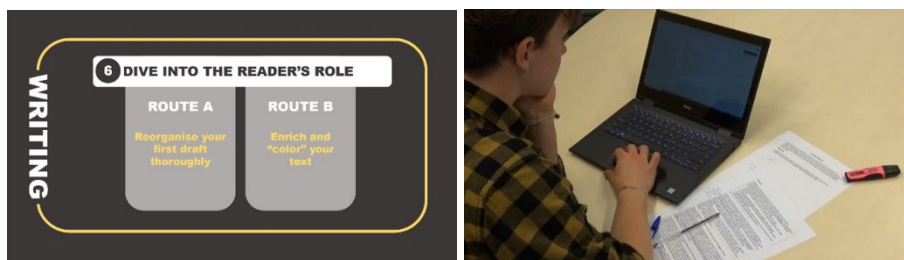


Figure 2: Stills from the Video with Modeling Peers (Translated; the Original Video Was in Dutch)

Table 1. Description of Instructional Stages and Learning Activities in the Design as Constructed

Session	Instructional stage	Learning activities	Description
Writing task (T1)	Develop background knowledge	Task experience	Students "experience" the whole task; they perform a first evaluative task. Prompt: To what extent are humans capable of true altruism?
Lesson 1	Describe it	Reflecting on the experience	The students write down what they thought was easy while performing the task and what was difficult for them.
		Building new knowledge Comparing the strategy to one's own experience	The teacher presents and explains the Read-Think-Write strategy. Students compare their task experience to the strategy presented.
	Model it (process)	Observing the strategy demonstrated by a modeling peer (video) Relating to the performance of others	Students watch a video together on the main screen. This video shows modeling peers, who demonstrate how each step of the strategy can be performed. As a processing activity, the students individually note the elements of the strategy that are most useful to them.
Lesson 2	Model it (product)	Analyzing/assessing peers' texts Generating criteria Applying new learning	The teacher selects two-three texts written in response to T1 to use as exemplars. Students individually note positive and negative aspects of each exemplar. In a class discussion of the exemplars, students generate a list of criteria. Students apply the criteria to a text fragment of their choice by rewriting/revising that fragment.
Writing task (T2)	Support it	Scaffolded practice	Students perform T2, practicing the strategy. This practice is scaffolded with the support of the teacher and the written guide (strategy cheat sheet).
Lesson 3	Support it	Providing peer feedback	Students bring the text they wrote in response to T2 to class and exchange their texts in a group of four students. They provide each other with feedback.
Writing task (T3)	Independent practice	Individual practice	Individual practice with the help of the written guide.

The design-as-constructed includes two feedback lessons. In Lesson 2, teachers are instructed to select two-three student texts written in response to T1 to discuss in class, and in Lesson 3, students discuss the texts they wrote in response to T2 with their peers in small groups consisting of three-four students. Analyzing and discussing exemplars is a means of conveying teachers' tacit knowledge about what criteria actually mean (Polanyi, 1973). In previous research (Orsmond et al., 2002), the discussion of exemplars has been proven to reduce differences between students' and teachers' assessments.

While completing T2, students can collaborate with their peers (Corcelles-Seuba & Castelló, 2015) and are supported by their teacher. At T3, students write independently, supported only by a written guide (the "strategy cheat sheet") (Martínez et al., 2015).

3.3 Teacher Guidance Activities

We designed a teacher guidance program (Figure 3) that fit the design requirements following the prestudy and the recommendations for professional development for teachers (Desimone, 2009; Van Veen et al., 2012). The program was designed to feature two objectives: teacher support and long-term teacher development.

The first objective was to provide support for the implementation of the instructional unit in two sessions. Session 1 (a one-on-one briefing) was included for teachers to take note of design principles for effective writing-to-learn tasks and to understand the underlying rationale of the instructional unit, ensuring content focus. Subsequently, teachers designed their own writing tasks to use in class. The research team provided feedback when requested. Teacher involvement in the design process enhanced active learning and coherence with teachers' goals.

Next, teachers implemented the instructional unit and reflected on their implementation process in an evaluative interview (Session 2). Since the activities were spread over one semester, the duration requirement was met. Although collaborative participation is a fifth critical feature of effective professional development, we considered an individual trajectory to be appropriate at this stage focusing on the beginning of innovation.

Our second objective was long-term teacher development. Therefore, we aimed to stimulate teachers' reflection. We included activities aimed at prompting contemplation of writing instruction and writing quality. We asked teachers to select for each task a benchmark text that they considered to be average in their group. This activity aimed to uncover teachers' expected level of progress at the group and student levels. In Session 3, teachers were asked to reflect on students' actual level of performance as determined by jury teams who assessed students' texts.

4. Evaluation and reflection phase

4.1 Research Design

Three philosophy teachers implemented the instructional unit in their 10th grade classes to develop students' philosophical writing. We investigated teachers' interaction with the design elements (tasks and instruction) by monitoring and evaluating the implementation process with the goal of answering RQ3-4.

To explore the effectiveness of the intervention for student learning, we used a semi-experimental design with three measurement occasions (T1-T2-T3). To estimate the impact of an instructional unit on three intact groups, an empirical intervention study was conducted. Independent jury teams assessed students' texts and teachers' tasks. Subsequently, teachers were asked to contemplate students' text scores. The results of the study are analyzed and presented as a multiple case study (Yin, 1994) using the teachers as cases.

4.2 Participants

Three qualified philosophy teachers (Teachers A, B and C) from three different schools participated in this study, each focusing on their own preuniversity 10th grade class (Groups A, B, and C, respectively). The teachers also participated in the prestudy and were committed to the development of students' disciplinary writing. The teachers had between six (Teacher B and C) and eight (Teacher A) years of experience teaching philosophy.

In total, fifty-six students (age: 16-17) participated (group sizes: A 22; B 17; C 17). Students actively consented to participate in this study; one student from Group C objected. Students' parents were informed of the study via regular communication channels; no parent objected. For all students, philosophy was a subject of choice. Students in Group B and C started philosophy in 10th grade. Students in Group A were taught philosophy in 7th and 8th grade as a mandatory subject. Students at School A might therefore be expected to have a more extensive knowledge base.

4.3 Procedures and Data Sources

Data sources were interwoven with procedures to guide teachers' implementation. Figure 3 provides an overview of implementation activities, research activities and data sources.

Teacher guidance activities	Research activities	Data sources	RQ
	Prestudy: Exploration of context	➤ Explorative interviews	RQ1-2
	Designing the instructional unit		
SESSION 1 Briefing	Preparing teachers for implementation		RQ3
Writing task design	Providing feedback	➤ Teacher-designed writing tasks	
Implementation of instructional unit	Check for implementation fidelity	➤ Classroom observations ➤ Teacher logs	
SESSION 2 Evaluative interview	Implementation assessment with a. teachers and b. students	➤ Evaluative interviews with teachers ➤ Evaluative interviews with students	
Defining the expected level	Teachers select benchmarks	➤ Selected benchmark texts ➤ Teachers' explanations	RQ4
External assessment of tasks and actual level of student achievement	Jury teams assess tasks and texts	➤ Field notes of jury team sessions ➤ Text scores	
SESSION 3 Reflective interview	Evaluation of student progress	➤ Reflective interviews with teachers	

Figure 3. Overview of Teacher Guidance Activities, Research Activities and Data Sources

4.3.1 Session 1: Briefing

With each teacher, we organized a one-hour, one-on-one briefing. We highlighted the essential elements of the intervention and explained the rationales underlying the design in general. We hypothesized that this briefing would improve the quality of the implementation. Subsequently, each teacher was instructed to design two writing tasks tailored to their regular year plan based on our design principles.

4.3.2 Writing tasks

As measurements for students' writing proficiency, we used three writing tasks (T1, T2, and T3) that were fully integrated into the instructional unit to avoid spending lesson time solely on research purposes. T1 was developed by the research team; this task was the same for all groups and functioned as the foundation of the strategy instruction. T1 asked students to discuss the concept of altruism ("To what extent does true altruism exist?") and contained four source texts (mean length: 117 words). The writing tasks for T2 and T3 differed from group to group since these tasks were designed by the participating teachers.

Students wrote their texts for T1, T2, and T3 on a computer in a computer room (Group A), using iPads (Group B), or laptops (Group C). Students' texts were rated by jury teams (section 4.3.5) and text length was determined; productivity was considered to be an additional indicator of text quality (Ferrari et al., 1998).

4.3.3 Session 2: Evaluative Interviews

Immediately after the intervention, we conducted evaluative interviews with teachers and students to obtain insights into their interaction with the design and implementation quality. Interviews with students were conducted in small groups (A: 5 students; B: 3 students; C: 2 students) to obtain information regarding the fidelity and validity of the intervention materials as well as students' self-perceived progress in writing. All interviews in this study took ca. one hour each, were audio-recorded and subsequently transcribed into written protocols. Interview guides are presented in Appendix A.

4.3.4 Benchmark Selection

After the implementation of the instructional unit, we asked each teacher to select for each assignment (T1, T2, and T3) one text from the written texts in their class that they would label as a benchmark text: an "average" text representing the average philosophical writing proficiency of students in the group. We asked for explanations of the strengths and weaknesses of each benchmark text and suggestions for improvement of the text. Teachers performed this task individually and communicated their findings via e-mail with the researcher. This activity was

aimed at encouraging teachers to contemplate the criteria they used for text quality. From a research perspective, this activity provided us with insights into teachers' expected level of student achievement.

4.3.5 Jury Team Sessions

Independent jury teams analyzed the texts written by students at T1, T2 and T3 to determine students' actual level of writing proficiency. The selected benchmark texts functioned as points of reference at a score of 100 points (Figure 4). Per assignment, all other texts were assigned points in comparison to this benchmark.

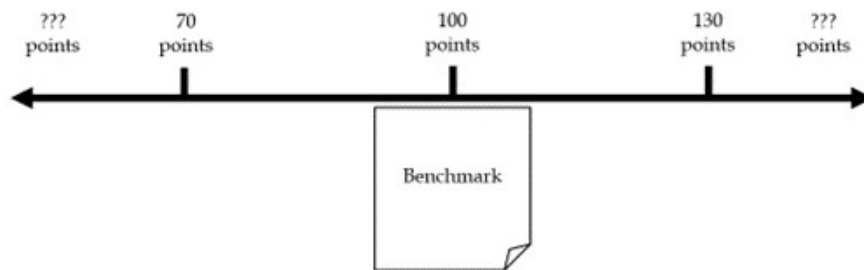


Figure 4: Assessment Scale

Jury teams consisted of one (Group A) or two (Group B, C) philosophy teachers in addition to a researcher (first author) who participated in all three of the teams. Each jury team collaboratively read and analyzed all three tasks of one group, discussed the benchmark texts and the accompanying explanations of the intervention teacher, and determined a score for three to four texts produced by students on every topic. The aim of the discussion was to obtain a shared understanding of the relevant criteria and standards. T1, T2 and T3 were discussed in random order. The collaborative sessions took 90 to 120 minutes each. The researcher took field notes during the sessions.

The jury team members individually assessed the remaining texts. Reliability coefficients for the individually assessed texts were .89 for the jury team for Group A, .91 for Group B, and .92 for Group C. Further analyses were conducted with mean scores.

4.3.6 Session 3: Reflective interviews

After the text rating procedure, we conducted reflective interviews with teachers. Each interview consisted of five parts, which were supported by visual representations of the group results. We started with a review of the designed writing tasks and the teacher's expectations regarding their students' results on

those tasks (Part 1), taking into account the level of the selected benchmark texts. Subsequently, the results for each task were presented in three graphs, one for each task; each graph contained all students' individual results for that particular task (Part 2). Then, we presented the group results in the form of an overall line graph showing group progress (Part 3) as well as in a quadrant visualizing individual students' progress from T1 to T3 (Part 4). Finally, we discussed the students whose results were atypical (Part 5): we showed the results at T1, T2, and T3 for a selection of students whose scores highly increased (+1 SD or +2 SD) or highly decreased (-1 SD or -2 SD) from T1 to T3. For each graph, we asked teachers (a) whether the results were consistent with what they had expected at the group and/or individual level; (b) to explore possible explanations for the results; and (c) to respond to trends and abnormalities at the student level.

For the analyses of the interviews, themes were extrapolated. For RQ3, we distinguished among four themes: teachers' interactions with (1) writing task design, (2) the strategy, (3) the modeling video, and (4) the discussion of exemplars. For RQ4, five themes were differentiated: (1) the characteristics of philosophical writing, (2) the evaluation of designed writing tasks, (3) benchmark quality, (4) expected progress, and (5) teacher perception of text quality. Per theme, all relevant data sources were involved in the elaboration (Figure 3).

5. Results

In the following sections, we present a cross-case analysis based on evaluative interviews, jury team discussions and reflective interviews with the goal of providing a more general perspective on RQ3 and RQ4. Using the technique of pattern matching (Yin, 1994), we compared the patterns of the separate cases to each other, and we compared these cross-case patterns to our predictions.

5.1 RQ3: Teachers' Interaction with the Design

5.1.1 RQ3a: Writing Task Design

Teachers B and C both designed two writing tasks that fully met the design principles. Teacher A was forced to adapt the principle of conciseness at T3; due to an unexpected shortage of the intervention period, T3 was performed in a test setting. Students received grades for their performance, which represented their final philosophy mark for the school year. The book chapter was handed out in advance: students were encouraged to read the source texts ahead of time independently to prepare for the exam.

The writing prompts that teachers formulated are presented in Table 2 (for elaborations, see Box 1).

Table 2. Overview of Writing Tasks

Group	Domain	T	Prompt	No. of sources	Total words (M)
A	Social/political philosophy	T2	Were Leopold and Loeb fully responsible for their actions, or did they have free will from which their actions sprang?	5	825 (165)
		T3	What does decision-making look like, ideally?	1*	8372
B	Philosophical anthropology	T2	To what extent are humans defined by 'reason' when compared to animals?	4	391 (98)
		T3	To what extent should humans allow emotions?	3	410 (137)
C	Philosophical anthropology	T2	To what extent is different treatment of men and women desirable in our society?	4	526 (131)
		T3	Should teachers at your school impose fewer obligations and rely more heavily on students' own responsibility?	4	474 (119)

* A book chapter was used as a source. This chapter contained different philosophers' views on the same topic. The students read the book chapter as a preparation before class, i.e., students had more time to write during class.

Box 1: Teacher-Designed Writing Tasks

Teacher A

Teacher A developed two tasks pertaining to social/political philosophy. For T2, Teacher A wanted students to discuss the issue of "free will". In the introduction to T2, the teacher described the famous case of Leopold and Loeb¹ (355 words). This description was followed by the following prompt: **Were Leopold and Loeb fully responsible for their actions, or did they have free will from which their actions sprang?** Four source texts were provided, which were all excerpts from the same book. In the first source (178 words), the concept of "insanity" was defined. The second text (403 words) described the history of Richard Kuklinski, who could not avoid becoming a hitman because of his life circumstances. The remaining three texts each defined concepts connected to the issue of whether free will exists: the principle of alternative possibilities (39 words), determinism (94 words), and libertarianism (111 words).

Group A's T3 asked the following question: **What does ideal decision-making look like?** Students were instructed to apply the question to a predefined current situation (note: at the time, Dutch farmers were protesting against the government because of governmental plans regarding the limitation of nitrogen emissions). Students were provided with one source text: a book chapter containing approx. 8300 words. In this chapter, four philosophers (Plato, Dahl, Mill, and Schumpeter) each explained their views on democracy. Students were instructed to refer to at least two of these philosophers' views.

Teacher B

Teacher B designed two tasks pertaining to philosophical anthropology. T2 was titled "the reasonable animal" and focused on the difference between animals and humans regarding reason. The task asked the following question: **To what extent are humans defined by reason?** Four source texts were provided. The first source (31 words) was a quotation from Aristotle stating that humans are the only creatures gifted with reason. The second source (74 words) was written by the philosopher Immanuel Kant, who argued that humans differ from things (including "mindless" animals) in rank and dignity. In the third source (178 words), Friedrich Nietzsche put man's role on earth into perspective. In the fourth source (98 words), the biologist Frans de Waal discussed why humans are inclined to downplay animals' intelligence.

T3 asked the following question: **To what extent should humans allow emotions?** Three sources were provided. The first source (185 words) was an excerpt from one of Seneca's letters to his friend Lucilius, in which he discussed stoicism. In the second source text (113 words), Aristotle argued that emotions should be allowed, as they can be purifying. In the third text (67 words), the contemporary philosopher Martha Nussbaum claimed that emotions can be a valuable source of knowledge.

Teacher C

Teacher C designed two tasks addressing philosophical anthropology. T2 focused on gender equality. The prompt was as follows: **To what extent is different treatment of men and women desirable in our society?** Four source texts were provided. The first text (157 words) was from philosopher Jean-Paul Sartre, who argued that we have no "inborn essence"; we are free to choose our identity. The second text (96 words) was written by Dick Swaab, a neurobiologist, who argued that men and women are naturally different from each other. The third source (111 words) was from a journalist who wrote a column in a newspaper about her observation that in kindergarten, children are already confronted with typical boys' and girls' toys. The fourth text (116 words) was written by a college student who indicated that medical tests are more focused on male bodies than on female bodies.

T3 focused on autonomy, applied to the students' own school. Students were presented the following statement: **Teachers of Edgewood High² should impose fewer obligations and rely more heavily on students' own responsibility.** Students were then asked to discuss different perspectives and weigh arguments. Four source texts were provided. In the first source text (127 words), philosopher Isaiah Berlin discussed the concepts of "positive" and "negative" freedom. In the second text (94 words), philosopher John Stuart Mill claimed that paternalism is allowed in some cases but not in others. In the third source (89 words), professor of developmental neuroscience Eveline Crone claimed that adolescents' brains are not yet mature. The final source text (100 words) was from Michel Foucault, who argued that freedom is unnecessarily restricted by institutions such as schools.

Teachers experienced difficulty selecting sources and delved into long source texts due to their desire to cover the whole issue. They wondered about the extent to which short writing tasks are suited for the development of philosophical writing. Nevertheless, teachers valued the short tasks, mainly due to their ability to help students develop the skill of writing a philosophical text in an efficient way. Furthermore, Teacher B noted writing in class made visible the fact that students differ greatly in terms of pace. Moreover, they noted, setting aside time for writing in class makes its importance more obvious.

Another challenge for teachers was the task of estimating the level of difficulty of primary texts that students could read and interpret individually. Although students were accustomed to discussing primary philosophical texts in class, they usually read those texts collaboratively.

Although they were not instructed to do so, teachers constructed tasks with increasing levels of difficulty. According to the teachers as well as the jury teams, task difficulty increased from T1 to T3. Difficulty-increasing factors included the

complexity of the issue, the number and complexity of the source texts, the coherence between the prompt and the sources, and the degree of coherence among the sources. Teachers agreed that T1 was of a "basic" level; the issue on which T1 focused was not particularly complex, contained a relatively high number of current sources, and included no texts from classic philosophers.

5.1.2 RQ3b: Writing Strategy Instruction

The strategy instruction consisted of three main elements, which are addressed below: the RTW strategy, the video with modeling peers, and feedback lessons.

RTW Strategy. Drawing attention to students' reading-writing process was a distinctive feature of the instructional design, which was indeed experienced as such by the teachers. Teacher A identified the process instruction using the RTW strategy as an addition to the writing instruction that they would usually provide, which they used to focus mainly on the product: the eventual text.

According to Teacher C, the strategy itself was perceived as "extensive" by students and therefore "impossible to perform during a 30-minute period". However, after the two practice tasks, Teacher C did observe improvement in students' ability to write a good text in 30 minutes. They included a similar writing task in the final exam for the school year, and they were amazed by students' performance. Students wrote up to 600-800 words, which Teacher C thought was "impressive".

Modeling Video. Teachers indicated the video stimulated students' self-reflective behavior. As teacher B explained, "The moment they saw other students demonstrating the strategy, they started to empathize. Do I recognize myself in this, do I do this too, (...) and which type of writer do I recognize myself in? It resulted in a good class discussion." Students and other teachers agreed with this claim; however, some students indicated that this situation caused them to feel unsure about their own approach when it deviated from the approaches taken by the students in the video.

Feedback Lessons. Teachers and students exhibited positive attitudes toward the feedback lessons. Overall, students were aware of the main goal underlying the exemplar discussions. As a student noted, "By reading someone else's text, you sometimes see what you can change about your own text" (C1).

However, teachers experienced it was difficult to direct students toward disciplinary criteria. Teachers noted that their group's list of criteria (which was, as instructed, the result of Lesson 2) was filled with generic criteria for high-quality texts; for example, "the text should contain coherence". The teachers questioned the relevance of philosophical writing; as Teacher C noted, "Students observe, for example, that a text has a lively introduction. Yes, okay, it does indeed have that, which is good, but do I really think that is important? It was very new to me, and I didn't really have a picture myself yet; I still don't". This quotation shows that

Teacher C would have preferred disciplinary criteria or at least a disciplinary focus for generic criteria; however, this tacit knowledge was difficult to bring to the surface.

Teacher B made an adjustment to the design-as-constructed in Lesson 2. They considered Group B to be quite a competitive group. Therefore, they were hesitant to select a weak and a strong example text; they thought that doing so would impact students' sense of safety. Instead, they chose to select example texts based on students' writing routines; they selected one text from a student whom they considered to be a free writer and one text from a student whom they suspected of having engaged in preplanning. Teacher B reported that because of the adjustment, the students were analyzing the texts rather than assessing them, which they believed to be favorable.

5.2 RQ4: Student Progress

During the reflective interviews, all teachers noted that they expected students to have improved their writing skills. Teachers A and C thought that this improvement would be visible in students' texts; however, Teacher B hesitated to endorse the same opinion. Nevertheless, Teacher B thought that students had improved in the sense that they had gained awareness of their processes.

From the benchmark selection procedure, one main quality criterion emerged: teachers wanted students to demonstrate their independent thinking. Teachers noted that their selected benchmarks exhibited increasingly independent thinking from T1 to T3. The independent jury teams agreed with this conclusion. Students' actual scores resulting from the text rating procedure and text length are presented in Table 3 and visualized in Figure 5.

Table 3. Mean Text Quality Scores and Number of Words

		T1	T2	T3	
		M (SD)	M (SD)	M (SD)	ES T1-T3
Text Quality Scores	A	86.1 (19.9)	83.1 (19.9)	99.9 (19.9)	.70
	B	78.1 (28.0)	77.1 (29.5)	80.9 (33.3)	.09
	C	92.6 (17.0)	92.0 (18.3)	97.2 (12.0)	.31
Productivity: Number of Words	A	336.7 (164.5)	274.9 (73.7)	530.3 (137.8)	1.28
	B	279.4 (87.0)	267.1 (120.2)	287.3 (109.6)	.08
	C	314.6 (71.4)	325.7 (119.1)	415.5 (113.6)	1.06

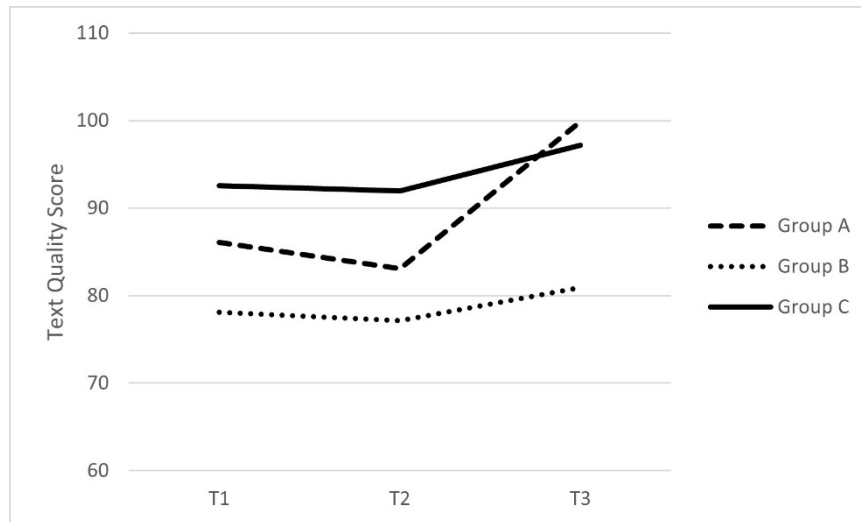


Figure 5: Mean Text Quality Scores at Each Measurement Occasion

At first glance, the flat lines shown in Figure 5 for Groups B and C may seem to indicate "no progress". However, the effect sizes for Groups A and C are not trivial: $d = .70$ in Group A and $d = .31$ in Group C (Table 3). When compared to the grade effect of a national baseline (Vandermeulen et al., 2023), the effect for Group A approximates a learning effect for writing of one school year ($d = .59$). Moreover, since benchmark text quality increased while the tasks became more complex, a flat line in Figure 5 is likely to indicate improvement in students' philosophical writing. However, it must be noted that the substantial increase from T2 to T3 in Group A might be due to the test setting and the extended writing time.

Overall, Group B scored lower than Groups A and C. The mean scores for this group were approximately 80 points, which was quite far from the benchmark score of 100. Group B also wrote the shortest texts on all measurement occasions. To explain this finding, Teacher B's view that philosophical thinking is difficult to exhibit in the context of short tasks might have played a role. As they stated, "I think students did not perform the tasks in a results-oriented way. And I'm afraid I teach this myself, in the sense that I don't pretend that students write such a text in three quarters of an hour impromptu. I think they have to think and weigh and reflect longer." As a result, the texts in Group B were unfinished, and incomplete texts were given low ratings by the jury team.

In all three groups, students' actual level was lower than expected by the teachers: teachers had expected that the average student would score 100, which was the level of the benchmark text. However, Group A and C could (nearly) achieve a mean score of 100 only at T3. Teachers thus consistently overestimated their students' progress.

Regarding productivity, text length and text quality scores were correlated: the longer the text, the higher the quality (T1: $r = .607$, $p < .001$; T2: $r = .793$, $p < .001$; T3: $r = .684$, $p < .001$). At T3, Group A wrote the longest texts, which was likely due to the fact that Group A had more time to write at T3, since source materials were read ahead of time. In Group C, students also wrote longer texts at T3 than at T1, while the length of the texts produced by Group B remained equal. The restricted time might have played a role in this context. As Teacher C noted, "Maybe this is just what a good writer can achieve within the timeframe provided". We thus conclude that the average student in Groups A and C wrote better texts at T3 than at T1 but that teachers estimated that students' average would be higher.

To understand what student progress in philosophical writing entails, we might analyze students' texts in more depth. Therefore, Example A and B present Group C's benchmark texts for T1 and T3, including Teacher C's explanation. Further to this, each Example presents a text from example student Luke, who showed a steady development in his philosophical writing. At T1, Luke wrote a text that was shorter and less elaborated than the benchmark text. Luke included source information in his text, but the last two sources are underrepresented, and a discussion or consideration of the arguments is lacking. The text fails to provide Luke's own distinctive perspective and independent thinking. Therefore, it received 82 points from the jury team. At T3 (Box 3), Luke wrote a text that was longer and more elaborated than the T3 benchmark. Luke's T3 text presents relevant source information in a structured way and relates the sources to the issue presented, involving his own school situation. The discussion of different theories and perspectives led Luke to a conclusion, showing he has given further thought to the issue. Therefore, this text obtained a score of 110 points; 10 points above the benchmark level.

Example A1. Benchmark Text at T1

Benchmark text (T1):

To what extent are humans capable of pure altruism?

Opinions differ on this, with the 1st half thinking that people do good deeds for their own sake and the other half thinking that people like to help others because they are born with those feelings.

Arguments from people who think people do good deeds out of selfishness:

Rose Spark thinks that people do good deeds for selfish feelings, consequences. For example, he/she says that one gets a good feeling, a warm glow, from helping others or avoiding guilt. Furthermore, help is reciprocal, and the other person will help you later.

Matt Ridley also believes that people do good deeds out of selfishness and agrees with Richard Dawkins that people want to preserve their genes. This explains why people care for children, they don't earn anything from it as individuals, but their genes are maintained. It also explains why people leave money as inheritance to their families.

Arguments from people who believe that people do not do good deeds out of selfishness:

Adam Smith writes in his work "The theory of moral sentiments" about that while humans do have selfishness, they also have principles by which they take an interest in others, for example compassion. Everyone also has empathy for other people; when someone else is sad you become sad yourself, and when someone else is happy you also become happy.

Matthieu Ricard believes that people have something in their DNA that makes them want to help others. This has been proven by several studies, which conclude that children already want to help others without social pressure. They do not need to learn this from their parents, for example, but do it themselves as early as 14 months of age. It was also concluded in an experiment that small children became happier from giving a candy to another person than getting a candy themselves.

You can conclude from all these arguments from both sides that this issue is a lot about your perspective on things. You may think that people will do anything to make themselves feel better or that people already naturally want to help others. Both are technically correct, and your opinion depends on how you look at it yourself.

Teacher C's explanation for choosing this text as a benchmark (100 points):

This text is mainly a summary of the sources, which are then outlined and linked to the given issue. The arguments are divided into two groups: pro and con. The conclusion is entirely relativistic. Most students seem to be able to read the sources and extract information that can be used to formulate an argument in response to the question.

However, the text lacks a consideration of the arguments, which is an essential part of philosophy, because only then independent thought emerges. By concluding that the answer depends on one's own perspective, the student in fact avoids answering the question.

Example A2. Luke's Text in Response to T1

In response to T1, Luke wrote:

Pure altruism, selflessly helping others without expecting anything in return, does it actually exist? According to social philosophy professor Roos Vonk, there is often a selfish reason behind a selfless act and the two concepts are not as opposite as many people think. Vonk suggests that there are all sorts of situations that at first glance appear to have only altruistic motives, only to find out later that there may also be selfish reasons behind them. For example, consider a situation where you do something for someone and expect them to do something in return for you.

According to evolutionary biologist Richard Dawkins, human morality can be explained by our "selfish" genes. By evolutionary theory, humans are not necessarily focused on themselves, but to maintain our genes. Science journalist Matt Ridley agrees with Dawkins, citing the example of leaving money to your children. Completely selfish people would not do this, but Dawkins' theory could explain it.

In contrast, moral philosopher Adam Smith and Buddhist monk Matthieu Ricard say that pure altruistic decisions can be made by humans. Smith uses the example of emotions that can pass from person to another person. Ricard even says that children are naturally born altruistic.

Note: Luke scored 82 points for this text, which was below the benchmark score.

Example B1. Benchmark Text at T3

Benchmark text (T3):

At a high school, many students and teachers are used to how the school system works. If you don't do your homework or if you are late, for example, you need to get a late pass and after several tardies you will get detention. But is this really the best system? Should teachers impose fewer obligations and rely more on students' own responsibility. I think so.

I think students need more responsibility at school because Eveline Crone's research shows that the brains of students in the pre-vocational track seem to develop faster than, say, students in the pre-university track. The scientists explain this by stating that students in pre-vocational education finish school at a younger age and therefore they have to deal with the demands of adult society earlier. If we, students at Edgewood High, are given more responsibility over our homework, for example, then their brain develops faster.

I also think that students need more responsibility at school, because positive freedom, which is now at school, can lead to paternalism, as Berlin says. I think paternalism should usually not be allowed in schools, so students therefore have more responsibility. But like John Stuart Mill, I think paternalism should only be allowed if it really harms someone. So, if students abuse their responsibility, for example by not coming to class anymore or not even trying to get good grades and thus harming their future, I think the school should be allowed to intervene.

So, I think that teachers at Edgewood High should impose fewer obligations and assume more of the students' own responsibility, because it means students' brains will develop faster. However, I also think that paternalism should sometimes be allowed at school, so that students cannot abuse their self-responsibility.

Teacher C's explanation for choosing this text as a benchmark (100 points):

In this text, the student has incorporated the arguments from the sources into his own views. However, real weighing of arguments does not occur. In a sense, this is not much more than setting out and applying an argument to the case, adding 'and I agree with that', but the student also manages to make the link between Berlin and Mill, which is interesting.

What this student could have done better, and which does emerge in texts by strong students, for example, is explicitly naming implications of a position. For example, in another student's conclusion, I read positive implications of imposing obligations, making abstract theory concrete and making the argument more persuasive. I would like to see the same with this student.

Example B2. Luke's Text in Response to T3

In response to T3, Luke wrote:

Fewer obligations for children?

At an average school, such as Edgewood High, young people are imposed an awful lot of obligations in the form of rules or standards. Students do not always agree with these obligations and sometimes they do not even know if they are useful at all. This raises the following question: Should teachers at Edgewood High impose fewer obligations and assume more of their own responsibility?

When we talk about freedom and obligations, we must first distinguish between two important types of freedom: positive freedom and negative freedom. According to Berlin, positive freedom occurs when a person feels the head of himself, and he makes his own choices. According to him [Berlin], paternalism, even when good intentions are involved, can be dangerous for someone and he would therefore argue that fewer obligations at Edgewood High could be a good idea.

According to some philosophers, such as John Stuart Mill, paternalism is something that should be used as little as possible. Nevertheless, if we can use measures to prevent people getting harmed we should do so. At Edgewood High, in some cases, obligations can have a positive impact, such as requiring younger students to stay in the schoolyard. Here, their freedom is taken away during school hours to ensure that they are safe. Moreover, there may be students among them who are unable to make autonomous decisions because they are still too young.

Scientifically, the prefrontal cortex, the part in your brain that determines how to make wise choices, is far from mature for middle school children. According to Eveline Crone, it doesn't finish developing until a person is 24. With this, you might argue that students are not independent enough to remove some obligations, but according to the law, a person from the age of 18 is wise and independent enough to be able to do (almost) everything by himself so this is not necessarily true.

An argument for this question could be given by Michel Foucault who says that our freedom is limited by institutions such as schools. According to Foucault, obligations would not even be necessary because the students themselves keep an eye on each other.

In conclusion, I personally think that fewer obligations at Edgewood High are not so much necessary because I agree with the words of Eveline Crone, an adolescent's brain is far from fully mature and therefore too much freedom could cause unwise choices.

Note: Luke scored 110 points for this text, which was above the benchmark score.

6. Discussion

With the aim of fostering secondary school students' philosophical writing, we designed an instructional unit that was open to contextual modifications (RQ1-2). Based on design principles for effective writing tasks, three philosophy teachers (A, B, and C) designed innovative source-based writing tasks to use in their 10th grade classes. They implemented the tasks and provided discipline-specific writing strategy instruction, which was supported by guiding sessions stimulating reflection. We conducted evaluative interviews regarding contextual adaptations, and we explored teachers' and students' interaction with the instructional design to test its resilience in open, ecologically valid situations (RQ3).

The results indicated that the teachers integrated the design into their contexts within the design parameters. The main challenge for teachers resided in the discussion of exemplars, which required functional knowledge of the criteria associated with philosophical texts, a situation that they noted required effort. Although we conclude that the teacher guidance program as a whole was successful in promoting teacher awareness of disciplinary writing instruction, it was clear that philosophical text quality criteria remained underemphasized.

Furthermore, we aimed to investigate the effects of teachers' changed practice on students' philosophical writing development (RQ4). External jury teams assessed the developed writing tasks and students' texts. We conducted reflective interviews with teachers concerning the expected and actual learner effects.

The results showed indications of effectiveness for the development of students' philosophical writing. After the intervention, students in Groups A and C exhibited more independent philosophical thinking in their texts, while tasks were judged to be more complex at T3 than at T1.

Student progress was in line with teachers' expectations. Evidently, the concept of progress according to teachers involved students' improving ability to complete tasks of increasing complexity. However, teachers' expected level of student achievement was not fully met. This discrepancy between expected and observed results is consistent with previous research on the accuracy of teachers' judgment, which has shown that teachers tend to overestimate their students, particularly low-achieving students (Südkamp et al., 2012).

Since we did not measure teachers' change in beliefs directly, we cannot draw direct conclusions concerning the change in teachers' beliefs regarding writing tasks and instruction. However, indications of change were visible in the interviews since teachers contemplated their role in the context of in-class writing and scaffolding students' individual reading-writing process, thus indicating a student-oriented approach (Kember, 1997).

Four main issues highlighted by this study merit further discussion: teachers' perspectives on what the functions of writing could be (§6.1.1); what a high-quality

philosophical text actually entails (§6.1.2); the value of writing strategy instruction (§6.1.3); and professional development with regard to literacy teaching (§6.1.4).

6.1 Main Issues

6.1.1 The Functions of Writing in Philosophy Education

With short tasks, we aimed to have students write within sight of the teacher to facilitate process support. Next to that, we aimed for writing-to-learn, promoting independent philosophical thinking. However, teachers expressed doubts about the extent to which short tasks are actually suited for this purpose. The tradition of philosophical writing advocates "slowing down" and "chewing" on the matter. A restricted timeframe limits this, especially when the tasks are too long to complete within the given timeframe. With the stimulation of philosophical thinking as a goal of writing, the balance between the length of the task and the time given should be optimal. The task should provide room for discovery on the one hand and restrain discovery in favor of communicative goals (presenting the key issue in a short text) on the other hand (Baaijen & Galbraith, 2018).

Although it could be an alternative to abandon subject goals in short writing tasks and to simply view short and long tasks as tasks with different goals, teachers remained dedicated to subject goals, and they tried to combine both purposes. Teachers suggested several solutions to the controversy, for example, to extend writing time to two lessons, which can be viewed as an indication that the design principles had become successfully internalized at that point.

6.1.2 Disciplinary Text Quality Criteria

Discussions of exemplars were included in feedback lessons. Since exemplars have proven to be most useful when employed in a dialogic way, with teachers and students jointly establishing a list of criteria for high-quality texts (Carless & Chan, 2017), teachers played an important role in this learning activity. Although both students and teachers valued the discussion of exemplars highly, this component of the strategy instruction raised several issues.

The first issue that teachers highlighted was that students were capable of formulating criteria, but not always the criteria that the teachers considered to be the most relevant. Students often focused on generic criteria for text quality, such as "no spelling errors". Teacher-led discussions are thus crucial to guide students away from solely meeting standards in favor of a focus on more discipline-specific criteria (Handley & Williams, 2011).

A second challenge teachers encountered when discussing exemplars, was that they required criteria and standards that are clear to themselves to be able to discuss these with students. This did not appear to be self-evident. Teachers noted that they needed practice to clarify for themselves the requirements that they

actually would prioritize. Knowledge of core constructs of disciplinary literacy (Goldman et al., 2016), e.g., disciplinary discourse and language structures, would probably be helpful, and these constructs should have been included in the teacher guidance program.

A third issue was the concern that exemplars might be intimidating for some students, constrain their creativity, or result in copying. However, exemplars are not intended to be model texts; they do not demonstrate "how a student should write a text" per se but rather show how peers approached the same assignment, which is unlikely to be intimidating. Furthermore, exemplars might provide students with new rhetorical solutions or strategies, which enrich their linguistic repertoire. Moreover, in our instructional unit, students did not discuss exemplars of topics about which they could write in the future but only discussed exemplars of topics about which they had already written. New tasks always contained new topics, which complicated copying behavior. Nevertheless, given that creativity is an important aspect of philosophical writing, we might consider if and how exemplars can play a role in the development of students' creativity in writing.

6.1.3 The Value of Writing Strategy Instruction

Teachers seemed to be satisfied with the concept of process instruction in general. Foremost, the strategy instruction caused students to reflect on their process according to teachers' observations. Monitoring might be the most impactful component of the strategy; after all, it characterizes both expert readers and writers (Bråten & Strømsø, 2011; Ferrari et al., 1998).

However, whether students utilized the strategy or changed their strategy after receiving the instruction remains unclear in this study. In the evaluative interviews, eight out of eleven students reported "no major changes" in their approach. Nevertheless, aspects of the strategy were noted to have been adopted; the elements that students described as having changed in their approach might be indicators of change even if they themselves mentioned experiencing "no change".

6.1.4 Professional Development

The set of guiding activities formed a distinctive type of professional development (PD) that aimed to help teachers feel equipped to incorporate writing tasks and instruction into their future teaching. With our teacher guidance activities, we met most of the criteria of effective professional development recommended by Desimone (2009); the only difference was that instead of a collective approach, an individual approach was adopted. We sought to equip teachers to teach disciplinary writing in three ways. First, we provided them with ready-to-use instructional materials to facilitate strategy instruction to ensure coherence with teachers' prior knowledge. Second, we offered design principles for the creating of writing-to-learn tasks. Teachers thus learned by doing. Third, we organized individual guiding

activities to enhance teacher awareness of disciplinary writing instruction and philosophical text quality to promote content focus. The confrontation with students' results that occurred in the reflective interviews led to reflection on the teachers' perceptiveness. Furthermore, the trajectory spanned approximately six months, which was consistent with the criterion of an ideal duration of approximately one semester.

The program could be optimized by including functional knowledge of criteria for philosophical texts. What might perhaps improve teachers' literacy practice further is an exchange of thoughts on text quality with their fellow philosophy teachers (cf. Van Drie & Stoel, 2020) after having experienced the corresponding challenges and profits. This approach would include the fifth critical feature of PD. As philosophy teachers are often alone in schools (philosophy is a subject that is mostly taught by one teacher per school), they might benefit from a learning community that focuses on literacy teaching (Desimone, 2009; Van Veen et al., 2012).

6.2 Strengths, Limitations and Directions for Future Research

This research involved a small-scale study with three philosophy teachers with the goal of exploring their interaction with an instructional unit, thereby providing writing process instruction to enhance students' philosophical writing. Our qualitative, contextualized approach provided us with the opportunity to explore this topic in depth. The reflective interview method we used was innovative and insightful. Teachers were asked to reflect on group results, which were produced by an independent jury team, thus ensuring the student factors that normally influence assessment to mitigate completely. For teachers, this method was highly informative. The confrontation with the group results automatically resulted in reflection.

A fact that must be considered, however, is that the three participating teachers chose to participate in the study and were thus already inclined to improve their teaching and make efforts to accomplish this goal. That fact may well cast the findings of this study in a certain light.

Another limitation might be the procedure we used to measure students' progress, such that writing tasks differed from group to group in terms of measurements, which might be viewed as a threat to internal validity. However, the rationale underlying this procedure was to design an instructional unit that was open to contextual modifications. Providing teachers with the opportunity to design tasks that are tailored to their own curriculum would be a major boost for ecological validity. Furthermore, the writing task design was connected to teachers' expectations of students' performance level, which was within the scope of our study.

The results of our study have several implications for both theoretical understanding and educational practice. This study contributes to our

understanding of disciplinary literacy development in an understudied field: the subject of philosophy. For example, this research initiated a discussion about what philosophical writing is, what it could be, what teachers expect of their students regarding the writing of philosophical texts, and what they consider to constitute student progress.

Furthermore, we provided insights into teachers' handling of innovative learning materials as they enter a new domain. These insights revealed that teachers who are unfamiliar with providing writing instruction require guidance and practice in teaching writing processes and discussing text quality with students and that this guidance can be provided by an integrated program of PD activities that stimulate teacher reflection.

Regarding practical implications, two types of relevance (De Vries, 1984) can be observed in this study, as teachers changed their practice and challenged their beliefs and attitudes. As a result, student learning improved. Thus, this study shows that process instruction can be a valuable addition to philosophy education since the results indicated the enhancement of students' philosophical writing.

It would be valuable for future research to explore the extent to which a reflective approach influences teachers' practice in the long term, as this was not part of the current study. Furthermore, we might explore the effects of intervening with instructional units that focus on philosophical reading-writing processes on a larger scale, possibly by focusing on the effects of such an approach on the most relevant criteria for philosophical writing or by examining this approach in tandem with philosophical reasoning measures.

Acknowledgements

We would like to greatly thank the teachers who participated in this study for their time and willingness to explore new lesson materials. They accomplished this task with great commitment despite the constant time constraints that characterize our teaching jobs. We also thank all the students who participated in this study and the jury teams who analyzed the students' texts for their time and efforts.

Declaration of Conflicting Interests

The author(s) declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

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

Notes

¹The names "Leopold and Loeb" refer to Nathan Leopold and Richard Loeb, two wealthy university students who kidnapped and murdered 14-year-old Bobby Franks in Chicago, Illinois, United States, in May 1924. They committed the murder in hopes of demonstrating a superior intellect that entitled them to commit a "perfect crime" without facing any repercussions.

²Edgewood High is a fictional school name. In the original task, Teacher C referred to their own school.

References

- Braaksma, M. A., Rijlaarsdam, G., Van den Bergh, H., & van Hout-Wolters, B. H. M. (2004). Observational learning and its effects on the orchestration of writing processes. *Cognition and Instruction*, 22(1), 1-36. https://doi.org/10.1207/s1532690Xci2201_1
- Borren, M. (2012). *Handleiding filosofische leesvaardigheden* [Manual for philosophical reading]. Rijksuniversiteit Groningen.
- Bråten, I., & Strømsø, H. I. (2011). Measuring strategic processing when students read multiple texts. *Metacognition and Learning*, 6(2), 111-130. <https://doi.org/10.1007/s11409-011-9075-7>
- 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. <https://doi.org/10.1017/CBO9781139048224.017>
- Carless, D., & Chan, K. K. H. (2017). Managing dialogic use of exemplars. *Assessment & Evaluation in Higher Education*, 42(6), 930-941. <https://doi.org/10.1080/02602938.2016.1211246>
- Chen, Y. C., Park, S., & Hand, B. (2016). Examining the use of talk and writing for students' development of scientific conceptual knowledge through constructing and critiquing arguments. *Cognition and Instruction*, 34(2), 100-147. <https://doi.org/10.1080/07370008.2016.1145120>
- Concepción, D. W. (2004). Reading philosophy with background knowledge and metacognition. *Teaching Philosophy*, 27(4), 351-368. <https://doi.org/10.5840/teachphil200427443>
- Corcelles-Seuba, M., & Castelló, M. (2015). Learning philosophical thinking through collaborative writing in secondary education. *Journal of Writing Research*, 7(1), 157-200. <https://doi.org/10.17239/jowr-2015.07.01.07>
- 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., Monte-Sano, C., Felton, M., Croninger, R., Jackson, C., & Worland Piantadosi, 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>
- Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, 38(3), 181-199. <https://doi.org/10.3102/0013189X08331140>
- De Vries, G. (1984). *De ontwikkeling van wetenschap. Een inleiding in de wetenschapsfilosofie* [The development of science. An introduction to philosophy of science]. Wolters-Noordhoff.

- Elbow, P. (1973). **Writing without teachers**. Oxford: Oxford University Press.
- 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. (2009). Writing as discovery. Teaching and learning writing. BJEP Monograph Series II, 6. *British Journal of Educational Psychology*, 2(6), p. 5-26. <https://doi.org/10.1348/978185409X421129>
- Galbraith, D. & Baaijen, V. M. (2018). The work of writing: Raiding the inarticulate. *Educational Psychologist*, 53(4), 238-257. <https://doi.org/10.1080/00461520.2018.1505515>
- Goldman, S. R. (2012). Adolescent literacy: Learning and understanding content. *The Future of Children*, 22(2), 89-116. <https://doi.org/10.1353/foc.2012.0011>
- Goldman, S. R., Britt, M. A., Brown, W., Cribb, G., George, M. A., Greenleaf, C., Lee, C. D., Shanahan, C., & Project READI (2016). Disciplinary literacies and learning to read for understanding: A conceptual framework for disciplinary literacy. *Educational Psychologist*, 51(2), 219-246. <https://doi.org/10.1080/00461520.2016.1168741>
- 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. https://doi.org/10.1163/9789004270480_003
- 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., Liu, X., Aitken, A., Ng, C., Bartlett, B., Harris, K. R., & Holzapfel, J. (2018). Effectiveness of literacy programs balancing reading and writing instruction: A meta-analysis. *Reading Research Quarterly*, 53(3), 279-304. <https://doi.org/10.1002/rrq.194>
- Handley, K., & Williams, L. (2011). From copying to learning: Using exemplars to engage students with assessment criteria and feedback. *Assessment & Evaluation in Higher Education*, 36(1), 95-108. <https://doi.org/10.1080/02602930903201669>
- Harris, K., & Graham, S. (1996). **Making the writing process work: Strategies for composition and self-regulation** (2nd ed.). Cambridge: Brookline Books.
- Heller, R., & Greenleaf, C. L. (2007). **Literacy instruction in the content areas: Getting to the core of middle and high school improvement**. Alliance for Excellent Education.
- Ho, A., Watkins, D., & Kelly, M. (2001). The conceptual change approach to improving teaching and learning: An evaluation of a Hong Kong staff development programme. *Higher Education*, 42, 143-169. <https://doi.org/10.1023/A:1017546216800>
- Holdinga, C. C. (2023). **Disciplinary writing: Four empirical studies on historical and philosophical literacy**. Dissertation, University of Amsterdam.
- Holdinga, C. C., Van Drie, J. P., Janssen, T. M., Rijlaarsdam, G. C. W. (2023). Writing to learn history: An instructional design study. *L1-Educational Studies in Language and Literature*, 23(1), 1-44. <https://doi.org/10.21248/l1esll.2023.23.1.526>
- Kember, D. (1997). A reconceptualisation of the research into university academics' conceptions of teaching. *Learning and Instruction*, 7(3), 255-275. [https://doi.org/10.1016/S0959-4752\(96\)00028-X](https://doi.org/10.1016/S0959-4752(96)00028-X)
- 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>
- Koek, A. (2020). Filosofisch lezen [Philosophical reading]. In: D. Berendsen, N. Kienstra, K. Poortier, & F. Rombout (Eds.) *Filosofie op school. Handboek vakdidactiek filosofie* (pp. 225-235). Amsterdam: Boom.

- Marsman, P. (2010). **Vakdossier filosofie** [Subject file philosophy]. Enschede: Stichting leerplanontwikkeling (SLO).
- Martínez, I., Mateos, 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>
- McDermott, M., & Hand, B. (2013). The impact of embedding multiple modes of representation within writing tasks on high school students' chemistry understanding. *Instructional Science*, 41(1), 217-246. <https://doi.org/10.1007/s11251-012-9225-6>
- McKenney, S., & Reeves, T. C. (2019). **Conducting educational design research**. Second Edition. Oxon/New York: Routledge. <https://doi.org/10.4324/9781315105642>
- 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., & 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>
- Monte-Sano, C., De La Paz, S., & Felton, M. (2014). Implementing a disciplinary-literacy curriculum for US history: learning from expert middle school teachers in diverse classrooms. *Journal of Curriculum Studies*, 46(4), 540-575, <https://doi.org/10.1080/00220272.2014.904444>
- OECD (2018). **PISA 2018 Results: Combined executive summaries. Volume I, II & III**. https://www.oecd.org/pisa/Combined_Executive_Summaries_PISA_2018.pdf
- 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. <https://doi.org/10.1080/0260293022000001337>
- Polanyi, M. (1973). **Personal knowledge**. London: Routledge/Kegan Paul.
- 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.
- Rijlaarsdam, G. & Janssen, T. (1996). How do we evaluate the literature curriculum? About a social frame of reference. In E. Marum (Ed.), **Children and books in the modern world: Contemporary perspectives on literacy** (pp. 75-98). London: The Falmer Press.
- Samuelowicz, K., & Bain, J.D. (1992). Conceptions of teaching held by academic teachers. *Higher Education*, 24(1), 93-111. <https://doi.org/10.1007/BF00138620>
- Shanahan, T., & Shanahan, C. (2017). Disciplinary literacy: Just the facts. *Educational Leadership*, 74(5), 18-22.
- Shanahan, C., Shanahan, T., & Misischia, C. (2011). Analysis of expert readers in three disciplines: History, mathematics, and chemistry. *Journal of Literacy Research*, 43(4), 393-429. <https://doi.org/10.1177/1086296X11424071>
- Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*, 104(3), 743. <https://doi.org/10.1037/A0027627>
- Torrance, M., Thomas, G. V., & Robinson, E. J. (1994). The writing strategies of graduate students in the social sciences. *Higher Education*, 27, 379-392. <https://doi.org/10.1007/BF03179901>
- Vandermeulen, N., Van Steendam, E., De Maeyer, S., & Rijlaarsdam, G. (2023). Writing process feedback based on keystroke logging and comparison with exemplars: Effects on the quality and process of synthesis texts. *Written Communication*, 40(1), 90-144. <https://doi.org/10.1177/07410883221127998>

- Van den Broek, P., & Helder, A. (2017). Cognitive processes in discourse comprehension: Passive processes, reader-initiated processes, and evolving mental representations. *Discourse Processes*, 54(5-6), 360-372. <https://doi.org/10.1080/0163853X.2017.1306677>
- Van Drie, J., & Stoel, G. (2020). 'Maar wat vind je hier nu goed aan?' Samen redeneringen van leerlingen analyseren als vorm van professionalisering ['What do think is good about it?' Collaborative assessment of students' texts as a way of professionalization]. *VELON Tijdschrift voor Lerarenopleiders*, 41(2), 87-100.
- Van Veen, K., Zwart, R., & Meirink, J. (2012). What makes teacher professional development effective? A literature review. In: M. Kooy, M., & K. van Veen (Eds.). *Teacher learning that matters: International perspectives*. Vol. 62, Chapter, 3, pp. 3-22. New York: Routledge.
- Velema, F., & Groza, T. (2020). Het schrijven van een filosofisch essay [Writing a philosophical essay]. In: D. Berendsen, N. Kienstra, K. Poortier, & F. Rombout (Eds.) *Filosofie op school. Handboek vakdidactiek filosofie* (pp. 215-223). Amsterdam: Boom.
- Villanueva, M. G., & Hand, B. (2011). Science for all: Engaging students with special needs in and about science. *Learning Disabilities Research & Practice*, 26(4), 233-240. <https://doi.org/10.1111/j.1540-5826.2011.00344.x>
- 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>
- Yin, R. K., (1994). *Case study research design and methods: Applied social research and methods series*. Thousand Oaks, CA: Sage Publications Inc.

Appendix A: Interview Guides

Explorative Interviews with Teachers (Prestudy)

A. Analysis of the writing task

1. 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?
2. Who developed this task?
3. Could this task also be a part of an assessment?
4. What are the characteristics of the source texts? (primary/secondary, genre, length, difficulty)
5. What are the characteristics of the text students wrote? (length, genre, audience)
6. What is the learning goal of the task?
7. To what extent have students achieved this learning goal?
8. Were you satisfied with students' results?

B. Analysis of students' texts

1. What is it that makes the weak text so weak?
2. Although you judge the weak text as weak, is there anything good in it?
3. What is it that makes the good text so good?
4. What could be improved in the good text?
5. To what extent is this task discipline-specific? What is typical history in this task?

C. Support and feedback

1. What did you and the students do prior to task performance? (instruction, support)
2. 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)
3. 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)
4. In retrospect: what would you do differently prior, during or after task performance?
5. What might help the student who wrote the good text improve? (prior, during, after performance)
6. What might help the student who wrote the weak text improve? (prior, during, after performance)

D. Cognitive processes

1. What cognitive processes are involved in the task? Please write them on sticky notes.
 2. 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.
 3. Could the job be done any other way, different from how you have described the process until now?
 4. What is the main thing you would like to see your students do differently in their process?
 5. How could you take care of that?
- E. Writing beliefs
1. How important is it for your students to have good writing skills?

Evaluative Interviews with Teachers

A. Descriptive

- a. What did you do in the regular lessons? Please, provide a description of the learning activities during these lessons.
- b. How did you conduct the intervention lessons? Did you skip or add elements? Why?

B. Evaluative

- a. To what extent do you think the lesson materials are practical? What improvements do you see?
- b. To what extent do you think the lesson materials are useful for learning? What improvements do you see?
- c. Which elements would you reuse?
- d. How did students respond to the lessons? Do you think the learning goal of writing better historical texts was achieved?

Evaluative Interviews with Students

A. Program differentiation

1. Were the philosophy lessons any different than you were used to? What was different? Was that positive or negative?

B. Motivation

2. How motivated were you during the intervention lessons? Why?
3. How motivated were your classmates?

C. Evaluative

4. How did you experience performing evaluative tasks?
5. What did you think of the intervention lessons?
6. What did you learn from the intervention lessons?
7. What would you have wanted to learn in addition?
8. What did you think of the strategy?
9. Would you keep using the strategy in future assignments?
10. Which step in the strategy was most useful for you?
11. Did you use the cheat sheet while making the assignments?
12. What did you think of the video?
13. What did you think of the modeling peer in the video?
14. What did you think of the model texts?
15. What would you tell next years' students who are starting the intervention lessons?
16. What improvements for the intervention lessons would you suggest?