



UvA-DARE (Digital Academic Repository)

Disciplinary writing

Four empirical studies on historical and philosophical literacy

Holdinga, C.C.

Publication date

2023

[Link to publication](#)

Citation for published version (APA):

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

General rights

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

Disclaimer/Complaints regulations

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

CHAPTER 5

FOSTERING PHILOSOPHY TEACHERS' DISCIPLINARY WRITING PRACTICE: A MULTIPLE-CASE DESIGN STUDY*

In this design study, an instructional unit open to contextual modifications was designed with the aim of fostering secondary school students' philosophical writing. Since teachers' teaching practices have been found to be guided by their beliefs and conceptions of teaching, the implementation of the design was prepared and supported by guiding sessions that stimulated philosophy teachers' thinking about writing and writing support in their discipline.

Based on design principles for effective writing tasks, three philosophy teachers designed innovative source-based writing tasks to use in their 10th grade classes. They subsequently implemented the tasks and provided additional discipline-specific writing strategy instruction, which was designed by the research team based on previous writing research.

Our study focused on teachers' interactions with the instructional design with the goal of testing the resilience of the design in open, ecologically valid situations. Furthermore, we aimed to investigate whether the unit was effective for students' development of philosophical writing.

A multiple-case approach using predominantly qualitative measures suited our research aims. After teachers had implemented the instructional unit in their own context, we conducted evaluative interviews with teachers and students concerning contextual adaptations. An external jury analyzed students' texts to determine their actual learning achievements. Teachers' insights into student progress were obtained from reflective interviews that featured comparisons between the observed and expected results.

* Chapter 5 is based on: Holdinga, C. C., Van Drie, J. P., & Rijlaarsdam, G. C. W. (2023). Fostering philosophy teachers' disciplinary writing practice: A multiple-case design study. Under review.

The results showed that teachers integrated the design with their specific contexts within the design parameters and judged the design to be feasible, valid, and effective for students' philosophical writing development. The experience of writing task design and the implementation of writing strategy instruction drove teachers to question the roles that writing tasks might play in philosophy education, the actual requirements of high-quality philosophical writing, and the possible value of writing strategy instruction in their discipline. These contemplations indicate changes in belief, which are necessary for genuine improvement in teacher practice. We therefore conclude that the findings indicate encouraging results regarding the effectiveness of an instructional unit accompanied by teacher guidance sessions that can promote reflection on the development of students' writing.

1. INTRODUCTION

The reading and writing skills that are necessary for the twenty-first century have increased. Today, literacy requires the ability to use reading and writing to acquire knowledge, solve problems and make decisions academically, personally and professionally (Goldman, 2012). However, 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). The impoverishment of students' literacy skills is problematic: these skills are a prerequisite for independent participation in our highly literate society. Moreover, literacy skills are crucial for students' subject matter learning and cognitive development (Graham et al., 2020).

The common view of the development of students' literacy in recent educational research has been that it is the responsibility of every teacher, not merely that of teachers in the language department or L1 teachers. Researchers have also noted that developing literacy is not merely a matter of developing proficiency in general reading and writing practices regardless of content (Moje, 2008; Shanahan & Shanahan, 2008). Instead, literacy is now viewed as a crucial element of the task of supporting students in understanding and 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 level of reading and writing skills. Students are required to read and interpret complex primary philosophical texts with the aim of enabling them to answer philosophical questions. Reading philosophical texts is quite difficult: these texts contain unfamiliar vocabulary, abstract ideas, complexly organized writing, and unsettling views (Concepción, 2004). In their search for a satisfactory answer to a philosophical question, students often read multiple texts to compare different philosophical views. Ultimately, students adopt a reasoned position toward the issue at hand. To demonstrate their proficiency in philosophical reasoning, students can be asked to express the position they support in, for example, the form of a Socratic conversation or a written text.

These activities also rely heavily on students' literacy skills. It might therefore 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. Furthermore, only 25% of the respondents reported that they explicitly explained to their students how to read a primary philosophical text. We thus conclude that there are opportunities to improve educational practice regarding disciplinary literacy development. With this study, we aimed to take advantage of this opportunity by developing, implementing, and evaluating an instructional unit intended to enhance students' philosophical reading and writing.

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). Moreover, research conducted by Goldman and colleagues (2016) produced a framework featuring five core concepts on which disciplines differ: (1) epistemology; (2) ways of enquiry 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 to 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, this topic has been explored with regard to 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 lack of research, we developed an instructional unit, which was inspired by De La Paz's research on the development of historical writing and is tailored to the subject of philosophy, 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 situation 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, thus ultimately fostering student learning (Desimone, 2009). The professional development of teachers is thus key to improving student achievement.

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). Teachers who view teaching as learning-oriented are likely to include more project-based learning activities, in which context students are responsible for their own learning process. In this case, the role of the teacher is to encourage and monitor students' work. However, teachers who view teaching as content-oriented are more likely to focus on imparting knowledge. Ho and colleagues (2001) have recommended a developmental

approach that focuses on conceptual change; these authors argued that real improvement in teachers 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 contemplations of writing tasks, instruction, and support.

According to the philosopher of science De Vries (1984), research can be practically relevant in a technical way by generating new techniques and ways of doing; however, it can also be relevant in a cultural way by generating new ways of expressing, viewing and thinking. Both types of relevance are addressed in this study, as teachers changed their practice and challenged their beliefs and attitudes. 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. In guiding sessions, three teachers were prepared for writing task design and the integration of this form of instruction into their 10th grade classes.

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 as they worked to improve their students' disciplinary writing in response to philosophical issues. 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. The third research question was therefore as follows:

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 as follows:

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 obtain a profound understanding of teachers' practices and beliefs, 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

We explored the educational context in a prestudy 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 11 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. For the full instructions for the interviewees and the interview guide, see Appendix E. The results showed that teachers used tasks that were mostly aimed at stimulating philosophical thinking. For example, students were asked to write a philosophical essay or an elaboration of a thought experiment. 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 level; however, all teachers indicated that not all students reached the required level. They reported finding it difficult to guide these students in this regard. The provision of process instruction or process feedback was scarce. Teachers seemed to lack knowledge about 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.

2.1 Design Principles

In Chapter 3, we established design principles for literacy development in history. Based on the prestudy 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 content focus 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: an open question that provokes a process of thinking and that cannot be solved by empirical research or observation. 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 containing multiple views on the issue evoke more 'sourcing' and 'referencing' (Britt & Rouet, 2012).

Our prestudy showed that philosophical essay tasks in upper secondary grades generally result in long texts: a length of 800-1000 words is not uncommon. Nevertheless, for two reasons, we aimed to emphasize concise tasks, resulting in short texts. First, we aimed to ensure that students wrote within sight of the teacher to facilitate the provision of process support. With short tasks, this goal is feasible. Second, we aimed to enable students to master a procedure and to provide them with multiple opportunities to practice this procedure. Shorter tasks would be most efficient in this case; writing multiple long texts would be tiring and demotivating for students.

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 of writing development in general (Graham & Harris, 2017). In another study that focused on the field of history (Chapter 4), we concluded that strategy instruction is also effective for disciplinary writing. In addition, research (Kieft et al., 2007) has shown that adapting writing instruction to students' writing strategies is effective for writing-to-learn, which 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 taken into account. Accordingly, the unit should be open to contextual modifications. When the design is flexible with regard to adaptation to different topics and levels of performance, it drastically enhances the feasibility of this approach for teachers.

Finally, our prestudy showed that philosophy teachers did not have much experience in teaching writing. Therefore, 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 taking into account 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.

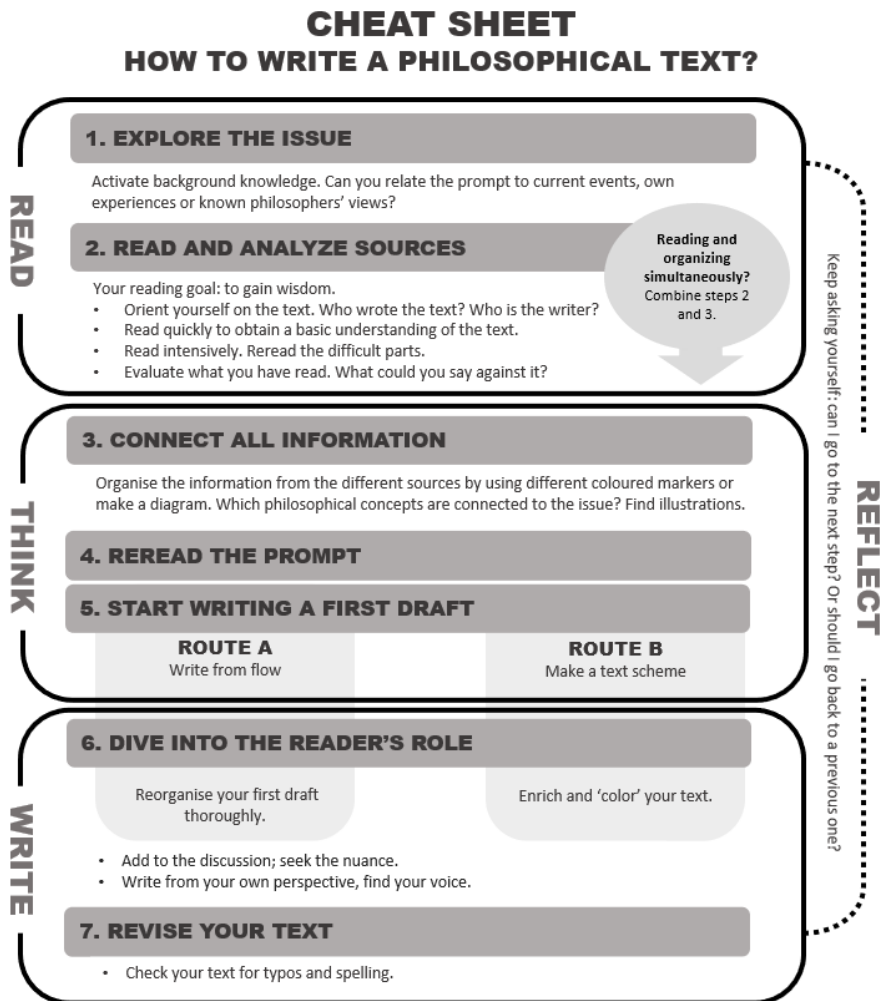
3.2 Writing Strategy Instruction

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

Based on a literature search, context exploration, and previous studies (Chapters 2, 3 and 4), we designed a genre- and discipline-specific Read-Think-Write strategy (RTW strategy). The RTW strategy is divided roughly into three main cognitive processes: reading, thinking, and writing. Furthermore, students are prompted to monitor the entire process since the three processes overlap and interact to a large extent. 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 5.1 shows the strategy as presented to students.

Figure 5.1. 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)



The first step in the seven-step strategy is to explore the issue presented and to activate prior knowledge (Corcelles Seuba & Castelló, 2015). Subsequently, students start reading the source materials to obtain enlightenment: to understand and reflect on what has been stated fully (Concepción, 2004) (Step 2). The phase of reading is aimed at obtaining a situation model: an elaborated

interpretation of the issue presented as described by the source (Britt & Rouet, 2012; Rouet & Britt, 2011). Creating such a model requires comprehension of content, reprocessing, recursive actions and a high standard of coherence (Van den Broek & Helder, 2007).

The next step is to connect the sources (Step 3) (Borren, 2012). This stage could partly overlap with the second step, as it might be possible for some students to connect information immediately with background knowledge or information they have previously read while reading. This step, which involves organizing source information, should lead to an integrated mental model: an internal representation of the issue discussed across texts (Britt & Rouet, 2012).

Step 4 is a reminder of monitoring. Subsequently, at the latest, students start writing at Step 5. The goal of writing during this phase is to stimulate knowledge constitution (Galbraith, 2009; Galbraith & Baaijen, 2018). The knowledge constitution process activates and articulates knowledge that is implicit or still diffuse. We offer two routes for writing that appeal to different writing preferences.

A first route is to write freely (Route A), i.e., using an intuitive style of writing. The free writer in fact follows Elbow's (1973) dual drafting strategy. In the first phase of writing (Route A, Step 5), the focus is on generating a first draft of the text without paying excessive attention to the rhetorical aspects that the text should or should not contain. In the second draft phase (Route A, Step 6), the writer focuses on the text itself, the language, the rhetoric, and the structure. Since we focus on short tasks, the drafts will not be actual drafts in the sense intended by Elbow. Instead, drafts progress in rapid succession.

A second route is preplanning (Route B), which involves a more decisive style of writing. Planning of the text (Route B, Step 5) might be performed at a local level - planning sentence by sentence - or at a global level - writing based on a global text plan. Creating a mind map or organizing information with the help of a text scheme is an alternative means of generating ideas. In the subsequent phase (Route B, Step 6), the text is expanded: students are instructed to "put flesh on the bones".

In general, Step 6 involves exploring the reader's perspective and considering the text critically. We then instruct students to organize the knowledge in a process of knowledge transformation (Galbraith & Baaijen, 2018). This process is necessary to transform students' knowledge into a rhetorically effective form. Finally, Step 7 focuses on editing.

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, which are described in Table 5.1.

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. The writing tasks were to be completed using a computer.

The instructional unit starts with a writing task (T1), and the students' goal is to "experience" the task. This experience functions as a "hook" for Lesson 1, when students reflect on their performance.

Subsequently, students are presented the Read-Think-Write strategy. The strategy is explained by the teacher and modeled by peers in a video as a mode of observational learning (Braaksma et al., 2004). During this video, explanatory parts are alternated with four modeling peers, who were 12th grade students from other schools (Figure 5.2).

Figure 5.2. Stills from the Video with Modeling Peers (Translated; the Original Video Was in Dutch)

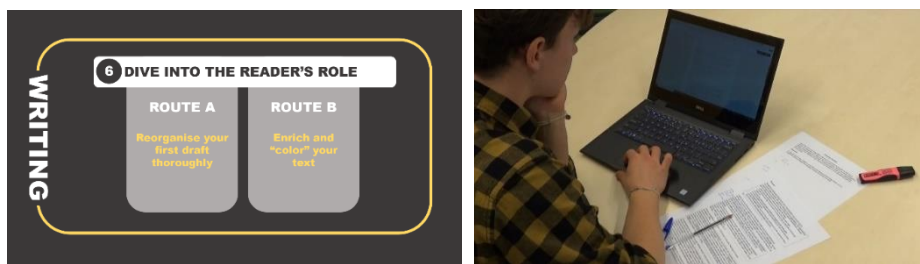


Table 5.1. Description of the Design-as-Constructed

Session	Stage	Learning activities	Description
Writing task (T1)	<i>Develop background knowledge</i>	Task experience	Students "experience" the whole task; they perform a first evaluative task.
Lesson 1	<i>Describe it</i>	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.
		<i>Model it (process)</i> 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	<i>Model it (product)</i>	Analyzing/assessing peers' texts	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.
		Generating criteria	In a class discussion of the exemplars, students generate a list of criteria.
		Applying new learning	Students apply the criteria to a text fragment of their choice by rewriting/revising that fragment.
Writing task (T2)	<i>Support it</i>	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	<i>Support it</i>	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)	<i>Independent practice</i>	Individual practice	Individual practice with the help of the written guide.

These peers were recruited to perform T1 on camera and filmed anonymously from an over-the-shoulder perspective. There was no script; students performed the task naturally in their preferred manner in approximately 40 minutes. One researcher (LH) selected scenes that were suitable for illustrating how to perform each step of the strategy and each "route", ultimately creating a video of 12 m 46 s in length. The illustrations contain scenes in which students struggle and engage in monitoring. For example, they conclude that they did not understand a source and start 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 are instructed to pause the video after each stage (read-think-write) and to reflect on the models in a class discussion.

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 that fit the design requirements following the prestudy and the recommendations for professional development for teachers (Desimone, 2009; Van Veen, et al., 2012). An overview of activities is presented in Figure 5.3.

The program was designed to feature two objectives. The first such objective is to support the integration and implementation of the instructional unit into teachers' practice. Initially, Session 1 was a one-on-one briefing (teacher and researcher) that enabled teachers to take note of design principles for effective writing-to-learn tasks and to understand the structure of the instructional unit and the underlying rationale. Based on this session, teachers could become well prepared for the implementation process in an authentic setting:

their own classroom. This session thus met the recommendation of 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. Subsequently, 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 goal was to equip teachers to implement high-quality writing instruction and support in their future teaching. 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 and RQ4.

To explore the effectiveness of the intervention for student learning, we used a semiexperimental design on three measurement occasions (T1-T2-T3); teacher-designed writing tasks functioned as measurement occasions at T2 and T3. This approach thus resulted in different measurements for each participating teacher at T2 and T3. T1 was similar for all groups and was developed by the research team.

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 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). All teachers also participated in the prestudy. The teachers participated voluntarily and were interested in the implementation of an intervention aimed at the development of students' disciplinary writing. The teachers had between six (Teacher B and C) and eight (Teacher A) years of experience teaching philosophy. All three teachers had obtained a master's degree in philosophy and had a teaching qualification (master's degree) for teaching philosophy in secondary education.

In Dutch secondary schools, philosophy is not a mandatory subject. At the schools of Teachers B and C, philosophy was taught beginning in 10th grade as a subject of choice. In Teacher A's school, philosophy was also taught in 7th and 8th grade as a mandatory subject for all students. Beginning in 10th grade, it was optional. All three groups thus consisted of students who chose philosophy as a subject; however, only students at School A might be expected to have a more extensive knowledge base.

In total, 56 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.

4.3 Procedures and Data Sources

Data sources were interwoven with procedures to guide teachers' implementation. An overview of implementation activities, research activities and data sources is provided in Figure 5.3.

Figure 5.3. Overview of Teacher Guidance Activities, Research Activities and Data Sources

Teacher guidance activities	Research activities	Data sources	RQ
	Prestudy: Exploration of context Designing the instructional unit	<ul style="list-style-type: none"> ➤ Explorative interviews 	RQ1-2
SESSION 1 Briefing Writing task design Implementation of instructional unit SESSION 2 Evaluative interview	Preparing teachers for implementation Providing feedback Check for implementation fidelity Implementation assessment with a. teachers and b. students	<ul style="list-style-type: none"> ➤ Teacher-designed writing tasks ➤ Classroom observations ➤ Teacher logs ➤ Evaluative interviews with teachers ➤ Evaluative interviews with students 	RQ3
Defining the expected level External assessment of tasks and actual level of student achievement SESSION 3 Reflective interview	Teachers select benchmarks Jury teams assess tasks and texts Evaluation of student progress	<ul style="list-style-type: none"> ➤ Selected benchmark texts ➤ Teachers' explanations ➤ Field notes of jury team sessions ➤ Text scores ➤ Reflective interviews with teachers 	RQ4

4.3.1 Session 1: Briefing

With each teacher, we organized a one-on-one session (lasting approximately one hour) to explain all aspects of the teaching manual. 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. Furthermore, it served as a baseline for assessing change over time. T1 did not rely excessively on philosophical concepts or an extensive philosophical knowledge base; it was intended for use in all participating groups and did not require prior philosophical knowledge.

T1 asked the following question: "To what extent are humans capable of true altruism?" Four source texts were provided (mean length: 117 words) to represent different perspectives on the issue. The first source text (123 words) was an excerpt from a blog written by social psychology professor Roos Vonk, who indicated that people may have all kinds of egoistic motives to help others. In the second text (106 words), journalist Matt Ridley discussed Richard Dawkins' theory of selfish genes: we would never have children if we did not have genes that were selfish. In the third text (113 words), philosopher Adam Smith claimed that even "the biggest villain" cares about other people's feelings. In the fourth source text (127 words), Buddhist writer and monk Matthieu Ricard cited a study proving that children act altruistically at a young age, thus suggesting that altruism could be "innate". The writing tasks for measurement occasions 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 on laptops (Group C).

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. Teacher interviews took approximately one hour. Interviews with students were conducted in small groups to obtain information regarding the fidelity and validity of the intervention materials as well as students' self-perceived progress in writing. Five students from school A (two female) were interviewed, as were three students from school B (all female) and three students from school C (one female). Student interviews took approximately 45 minutes. All interviews were audio-recorded and subsequently transcribed into written protocols. Interview guides are presented in Appendix E.

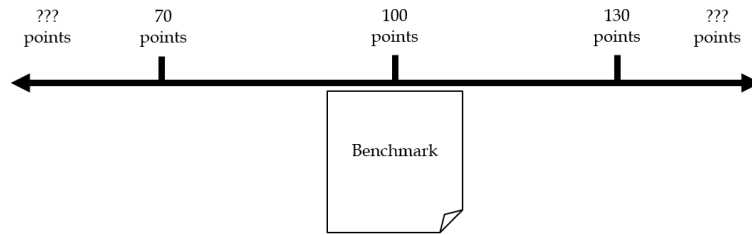
4.3.4 Benchmark Selection

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 invited teachers to elaborate on their decisions: 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. 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 5.4). Per assignment, all other texts were assigned points in comparison to this benchmark. For example, a text could score 80 points if it was judged to be of lesser quality or 120 points if its quality was rated higher.

Figure 5.4. Assessment Scale



Jury teams consisted of one (Group A) or two (Group B, C) philosophy teachers in addition to a researcher (LH) 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 approximately 90 to 120 minutes each. The researcher took field notes during the sessions.

The remaining texts were individually assessed by the jury team members. 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 by reference to mean scores.

4.3.6 Session 3: Reflective Interviews

After the text rating procedure, we conducted reflective interviews with Teachers A, B, and C. Each interview consisted of five parts, most of 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 either 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. The reflective interviews took approximately one hour, were audiotaped and were subsequently transcribed into written protocols.

Based on 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 (difficulty/length) of designed writing tasks, (3) benchmark quality, (4) expected progress, and (5) perceptiveness. Per theme, all relevant data sources were involved in the elaboration (Figure 5.3).

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 were instructed to design reading-writing tasks based on the principles of writing-to-learn. The prompts that teachers formulated are presented in Table 5.2 (for elaborations, see Box 1).

Teachers B and C designed writing tasks that fully met the design principles. With respect to T3, Teacher A adapted the principle of conciseness. This adaptation was a forced adjustment; due to an unexpected shortage during 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.

Table 5.2. Overview of Writing Tasks

Group	T	Domain	Prompt	No. of sources	Total words (M)
A	T2	Social/political philosophy	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	T2	Philosophical anthropology	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	T2	Philosophical anthropology	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 Loeb* (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).

* The 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.

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 with regard to 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 [school name] 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.

Overall, the guidelines to make each task performable within a period of 30-40 minutes turned out to be difficult. Teachers experienced difficulty selecting sources and delved into long source texts due to their desire to cover the

whole issue. This difficulty might have been caused by subject-related traditions: teachers of philosophy are accustomed to more extensive writing tasks, in which context students are regularly required to write at their own pace. However, in our study, we asked teachers to design shorter writing tasks because we wanted students to write within sight of their teacher, thereby enabling teachers to observe and support the process. The element of writing in class and within sight of the teacher was valued. Teacher B indicated that this approach was insightful; it 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.

Both students and teachers thus experienced the tasks as "different", not because of their evaluative nature but rather particularly due to their shortness. We might wonder about the extent to which short writing tasks are suited for the development of philosophical writing. The norm in philosophical writing tasks is that philosophical thinking is central. The thinking process is relevant and reflected in the text. Subject-specific genres, such as philosophical essays, even revolve entirely around thinking processes. In the short tasks used in our study, students seemed to adopt this default view of philosophical tasks: a transfer from one philosophical task to another. Teacher B indicated that if students had viewed the writing tasks as more similar to, for example, test questions, in which context time pressure is common, they would have approached the task in a more pragmatic and result-oriented manner. With respect to the short writing tasks, students did not seem so pragmatic: "Let's write a quick conclusion to it" was not an idea that occurred to them. Nevertheless, teachers valued the short tasks due to their ability to help students develop the skill of writing a philosophical text in an efficient way with the aim of mastering the procedure of reading and writing in the field of philosophy. To reflect the typical thinking and writing procedures used in philosophy more accurately, teachers suggested extending the time available for thinking, for example, by spreading the writing over two lessons or by allowing a second draft to be written after feedback lessons.

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. Teacher B indicated that they had underestimated how difficult it was for students to read primary texts individually. They wondered whether it would have been more

effective to use longer but more accessible texts and to provide more explanation and decreased information density instead of short but complex excerpts. 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 (1) the complexity of the issue in general (how abstract is the topic, whether the task contains philosophical concepts that are assumed to be familiar), (2) the number and complexity of the source texts (how many perspectives are to be included, whether any new concepts must be understood, whether the student has prior knowledge of the author/philosopher in question), (3) the coherence between the prompt and the sources (how the sources can be related to the issue), and (4) the degree of coherence among the sources (to what extent the sources can be related to each other). 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, was written by "new" philosophers, 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. Students received direct instruction by means of a "cheat sheet" that presented the 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. They were familiar with the approach of providing students direct instruction on text structure: for example, by identifying necessary elements in an introduction. Teacher A remarked that they would now know how to provide process instruction in the future because of the didactic tools with which they were provided.

Teacher B also valued the process instruction: when philosophy teachers paid attention to the processes associated with the development of philosophical skills, they thought that this approach would make the subject more "learnable" for students. Teacher B noted philosophy is commonly viewed by students as a subject that is mainly suitable for high achievers. However,

when philosophy teachers highlighted the development of processes, they encouraged students' growth mindset instead, thereby making the subject more attractive.

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. With regard to the video, Teacher B indicated that it was useful with regard to students' self-reflective behavior. As this teacher 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. The instructional unit contained two feedback lessons (Lesson 2 and 3), each of which followed a writing assignment (T1, T2). In general, teachers exhibited positive attitudes toward the feedback lessons. As Teacher A said, "Students discovered that others could discuss the same issue with a completely different approach. That was very enlightening." Indeed, this enlightenment concerning the available possibilities was one of the goals of presenting the example texts.

Teachers indicated that students were motivated to discuss the texts. The interviews with students confirmed this claim. Overall, students were aware of the main goal underlying the exemplar discussions. As one student (C1) noted, "By reading someone else's text, you sometimes see what you can change about your own text."

However, teachers also experienced difficulties. The main difficulty pertained to the task of directing students toward critical criteria for philosophical texts. Teacher B noted that Group B'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" and "the issue should

be introduced properly". Nevertheless, Teacher B was satisfied with the result. As this teacher noted, "The criteria list relied more on structure than on philosophical content, but I think that it creates space for students to internalize the content, and then they have more confidence in how they structure their text."

Teacher C also noted that students often suggested generic criteria. They found it difficult to respond to such suggestions because they questioned their relevance for philosophical writing: "Students suggested, for example, that this 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.

Another difficulty for teachers pertained to the need to continue linking the product (what is a good text) to the process (how to write a good text). Teacher C said, "From Lesson 2 onward, we deviated from the approach of how to tackle the task. We got more into the content, of what a good text should actually contain."

Teacher B made an adjustment to the design-as-constructed in Lesson 2, which was a feedback lesson. 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. Teacher B 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. During Lesson 2, Teacher B learned that the strategies they suspected the two students of having used were in fact not those that the students claimed to have used: "It turned out to be false, but that did not matter because it was visible that the students wrote from different angles." Teacher B reported that as a result of the adjustment, the students analyzed 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

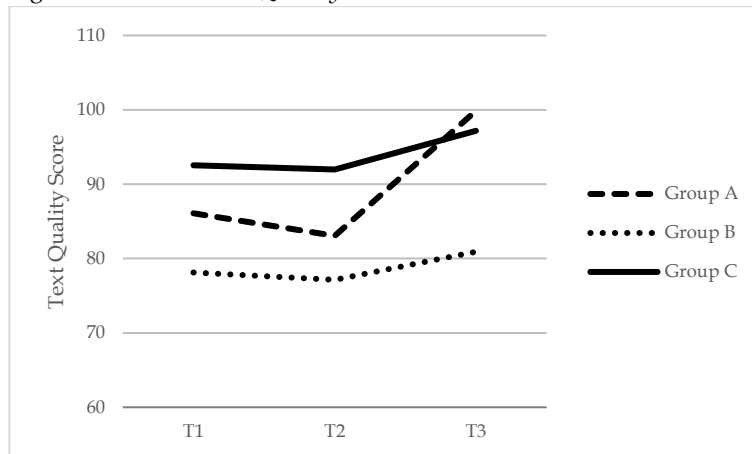
students had improved in the sense that they had gained awareness of their processes.

From the benchmark selection procedure, one main quality criterion emerged: teachers wanted students to demonstrate their independent thinking. Teachers indicated that it was most important for students to relate to the philosophical issue or source texts at hand and to consider different arguments and to develop a personal opinion that could indicate independent thinking. Teachers noted that their selected benchmarks exhibited increasingly independent thinking from T1 to T3. The jury teams agreed with this conclusion. Students' actual scores resulting from the text rating procedure and text length are presented in Table 5.3 and visualized in Figure 5.5.

Table 5.3. Mean Text Quality Scores and Number of Words

	Group	T1 M (SD)	T2 M (SD)	T3 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 5.5. Mean Text Quality Scores at Each Measurement Occasion



At first glance, the flat lines shown in Figure 5.5 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 5.3). 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 6 is likely to indicate improvement in students' philosophical writing.

As noted, Group A's scores increased substantially from T2 to T3. Teacher A thought that the test setting at T3 had positively influenced the results. Students were motivated to write good texts because they aspired to obtain good final grades. In addition, students had time to prepare the writing since they had received the source materials in advance. Teacher A noted that students who had read and analyzed the source materials actively and thoroughly during the class prior to the writing session also wrote better texts. Moreover, this teacher attributed the "drop" at T2 to the difficult class circumstances, which influenced students' motivation negatively.

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. Teacher B was surprised by this result, which caused this teacher to consider whether their expectations of students' performance were unrealistic. 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 this teacher 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. Furthermore, Teacher B indicated that their group was not very high-performing in general but that they had ignored that fact when selecting benchmark texts; their choice for the benchmark texts was based on what they would expect from the "average student" in a regular performing group.

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, only Group A was capable of achieving a mean score of 100 at T3, although Group C was close to this mark. Teachers thus consistently overestimated their students. This

discrepancy is in line 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).

Markedly, in none of the groups was progress visible from T1 to T2, while strategy instruction was provided between those measurement occasions. A plausible explanation for this finding might be the increase in task difficulty from T1 to T2. T2 and T3 were consistently assessed to be more complex and to contain more philosophical concepts. Moreover, T2 and T3 were embedded in content lessons, thus requiring a philosophical knowledge base. Intervention lessons were therefore interspersed with content lessons.

An alternative explanation for this finding might be that students received feedback only after writing T2. This feedback session was perceived as very instructive; however, any advantages from this session could not be reflected by students' scores for T2.

Productivity was considered to be an additional indicator of text quality. Previous research has shown that good writers write longer texts (Ferrari et al., 1998). In our study, this concept also seemed to be applicable, since 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; however, this major increase in Group A was likely due to the test setting. Moreover, since source materials were read ahead of time, students had more time to write, which is likely to result in longer texts. 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.

6. DISCUSSION

With the aim of fostering secondary school students' philosophical writing, we conducted a design study that consisted of three phases: design, implementation, and evaluation. In the first phase, 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 subsequently implemented the tasks and provided additional discipline-specific writing strategy instruction, which was designed by the research team (RQ2) based on previous writing research. Since it has been reported that teachers' teaching practices are guided by their beliefs and conceptions of teaching, the implementation was prepared and supported by guiding sessions that stimulated philosophy teachers' thinking about writing and writing support in their discipline.

After teachers had implemented the instructional unit in their specific context, we conducted evaluative interviews regarding contextual adaptations. We explored teachers' and students' interaction with the instructional design to test its resilience in open, ecologically valid situations (RQ3).

The results indicated that our design principles for effective writing tasks were feasible. The most challenging aspect of writing task design was to ensure that the tasks remained short. However, teachers did highlight opportunities to improve their tasks, which can be viewed as an indication that the design principles had become successfully internalized at that point. Regarding strategy instruction, the results showed that teachers integrated the design with their contexts within the design parameters and judged the design to be feasible. Teachers indicated that they felt supported in the task of providing process instruction. However, they also experienced that functional knowledge of the criteria associated with philosophical texts was required to provide feedback, a situation that they noted required effort. This aspect remained underemphasized in the teacher guidance program.

Furthermore, we aimed to investigate whether the unit was effective for 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.

We conclude that the design-as-constructed 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 the expectations of Teacher A and C; however, their expected level of student achievement was not fully met; teachers had estimated that students' writing would have improved despite the fact that the task difficulty had increased.

Group B exhibited no increase in writing quality: however, Teacher B reported that students had gained an awareness of their writing process, which they thought was valuable. Moreover, even a flat line from T1 to T3 might have indicated growth since students were able to continue at the same score level even when the task difficulty and benchmark level increased. The concept of progress according to teachers thus involved students' improving ability to complete tasks of increasing complexity.

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

In our study, we instructed teachers to design writing tasks that were suitable for a timeframe of 30-40 minutes. Our rationale for the use of short tasks instead of longer tasks was twofold: (1) we aimed to have students write within sight of the teacher to facilitate process support, and (2) we aimed to enable students to master a procedure and to provide opportunities for them to practice this procedure. In this case, shorter tasks would be most efficient.

Teachers and students experienced the nature of the writing tasks as new: they were used to more extensive writing tasks that featured extended writing time. Although they appreciated the benefit of efficiency with regard to instructing students in the procedures associated with philosophical reading and writing, they also wondered whether the tasks were appropriate to the stimulation of philosophical reasoning. The tradition of philosophical writing advocates "slowing down" and "chewing" on the matter. With short writing tasks, the benefit of "slowing down" diminishes, especially when the tasks are too long to complete within a given timeframe. The balance between the length of the task and the time given should be optimal. It was challenging for teachers to ensure that the size of the tasks remained sufficiently limited to

provide room for discovery on the one hand and to restrain discovery in favor of communicative goals (presenting the key issue in a short text) on the other hand (Baaijen & Galbraith, 2018). However, teachers valued the fact that students wrote within their sight, thus giving them the opportunity to support students' reading and writing process. This fact exposed differences in pace among students that these teachers had never previously considered.

Teachers identified several ways of improving their writing tasks to make them more appropriate to the discipline of philosophy: (1) writing tasks could be shortened (i.e., less sources) to mitigate the philosophical complexity and narrow students' focus solely on key issues; (2) writing could be spread over two lessons to extend the available writing time; or (3) the writing lessons could be viewed as "first draft" lessons, and after a feedback lesson, students could be provided with an opportunity to rewrite their drafts.

6.1.2 Disciplinary Text Quality Criteria

In the instructional unit under study, we included discussions of exemplars to improve students' writing performance 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 they considered to be the most relevant. Students, for example, focused on generic criteria for text quality, such as "a good structure", "clear formulations", or "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, which might be less superficial (Handley & Williams, 2011). For example, teachers considered it to be important that students' texts deliberated on different perspectives instead of merely summarizing them.

This issue of criteria generation remaining at a superficial level (e.g., structural or spelling issues) during class discussions raised another issue: teachers are required to have criteria and standards that are clear to themselves to be able to communicate or discuss these with students. Teachers must have the confidence and functional knowledge necessary to do so. However, this awareness 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 Goldman's (2016) core concepts of disciplinary literacy would probably be helpful, and this framework should have been included in Session 1.

Teacher C indicated that they felt more confident after several rounds of "practice" discussing exemplars: after the intervention, they felt that they knew the difficulties that students experienced during writing and the mistakes that were commonly made, which made them more confident. Guiding activities (benchmark selection, reflective interview) might also have offered additional value by bringing tacit knowledge to the surface. A related challenge for teachers was to continue establish a connection between product criteria and the process that was the subject of the instruction. However, in this case as well, practice might lead to more routine behavior.

A third issue was raised by Teacher B. They thought that exemplars might be intimidating for some students, constrain their creativity, or result in copying. Are these fears reasonable? Exemplars are not intended to be model texts; exemplars do not demonstrate "how a student should write a text" per se but rather show how a peer approached the same assignment. Exemplars are thus not perfect texts and are therefore unlikely to be intimidating. However, Teacher B's remark raises a question that was also posed by Handley and Williams (2011) regarding whether students consider good exemplars to be models and whether their availability facilitates copying. However, although exemplars might provide students with ideas, these new ideas are intended to expand their repertoire. This fact is not a limitation of creativity, but it contributes to seeing possible ways of elaborating a text and expanding one's linguistic repertoire. Moreover, in our instructional unit, students did not discuss exemplars of topics about which they could write in the future but rather discussed only exemplars of topics about which they had already written. New tasks always contained new topics, which complicated copying behavior. Elements that can be used in new texts might be rhetorical solutions or strategies, which might only enrich students' future writing.

6.1.3 The Value of Writing Strategy Instruction

None of the teachers in our study made adjustments to Lesson 1, which contained direct strategy instruction and modeling peers demonstrating the RTW strategy. One explanation for this fact might be that this lesson offered little room for adjustments. In general, teachers seemed to be satisfied with the strategy and with the concept of process instruction in general.

The question of whether students utilized the strategy or changed their strategy after receiving the instruction remains unclear in this study. In the evaluative interviews, most students reported "no major changes" in their approach. However, aspects of the strategy were noted to have been adopted, for example, ways of integrating sources into a text. Accordingly, the elements that students described as having changed in their approach might be indicators of change even if they themselves mentioned experiencing "no change".

Foremost, the strategy instruction caused students to reflect on their process according to teachers' observations. Monitoring is also part of the strategy itself; at Step 4 of the strategy, students are prompted to keep sight of what is actually asked. In addition to that explicit reference, students' attention was drawn to "monitoring" visually in the "cheat sheet". From this perspective, 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). Directing students' awareness toward this aspect of the writing process seemed to be a valuable addition to philosophical writing instruction.

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); only collective participation was absent. 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.

Based on the results, we conclude that the guiding activities as a whole were successful in promoting teacher awareness of disciplinary writing instruction and philosophical text quality. We equipped teachers with

knowledge related to writing tasks and with the didactic tools necessary to address and support writing processes.

The program could be optimized by including functional knowledge of criteria for philosophical texts, for example, by emphasizing the framework developed by Goldman (et al., 2016). 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 that the student factors that normally influence assessment to mitigated 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, which cognitive processes it might or should address, 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 at school. 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, we think that our study was significant in two ways. As the philosopher of science De Vries (1984) noted, research can be practically relevant in a technical way by generating new techniques and ways of doing; however, it can also be relevant in a cultural way by generating new ways of expressing, viewing and thinking. Both types of relevance 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 genuinely 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.