Effects of instruction in writing-to-learn in different disciplines and types of education

Published by
LOT
Binnengasthuisstraat 9
1012 ZA Amsterdam
The Netherlands

phone: +31 20 525 2461

e-mail: lot@uva.nl
http://www.lotschool.nl

Cover and illustrations: Geert Hermkens, Löss Ontwerp

ISBN: 978-94-6093-443-8
DOI: https://dx.medra.org/10.48273/LOT0658
NUR: 616

Effects of instruction in writing-to-learn in different disciplines and types of education

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor

aan de Universiteit van Amsterdam

op gezag van de Rector Magnificus

prof. dr. ir. P.P.C.C. Verbeek

ten overstaan van een door het College voor Promoties ingestelde ommissie,

in het openbaar te verdedigen in de Agnietenkapel

op donderdag 1 februari 2024, te 10.00 uur

door Aartje Cornelia van Dijk

geboren te Zeist

Promotiecommissie

Promotor:	prof. dr. F. Kuiken	Universiteit van Amsterdam
Copromotor:	dr. A.J.S. van Gelderen	Hogeschool Rotterdam
Overige leden:	prof. dr. S.J. Andringa	Universiteit van Amsterdam
	prof. dr. C.A.M. van Boxtel	Universiteit van Amsterdam
	prof. dr. C.M. de Glopper	Rijksuniversiteit Groningen
	prof. dr. J.J.M. Schoonen	Radboud Universiteit Nijmegen
	dr. P.J.F. Snellings	Universiteit van Amsterdam
	dr. J. Smit	Hogeschool Utrecht

Faculteit der Geesteswetenschappen

Het hier beschreven onderzoek en de uitgave van dit proefschrift werden mede mogelijk gemaakt door steun van Hogeschool Rotterdam.

Table of contents

Author Contributions

Chapter 1 General introduction
This chapter was written by Aartje van Dijk with feedback from Folkert Kuiken and Amos van Gelderen.

Chapter 2 Which types of instruction in writing-to-learn lead to insight and topic knowledge in different disciplines? A review of empirical studies
This chapter is a slightly modified version of the published study: Van Dijk, A., Van Gelderen, A., & Kuiken, F., (2022). Which types of instruction in wrting-to-learn lead to insight and topic knowledge in different disciplines? A review of empirical studies. *Review of Education*. https://doi.org/10.1002/rev3.3359

The study was designed in consultation with Amos van Gelderen and Folkert Kuiken. Aartje van Dijk collected relevant studies with feedback from Amos van Gelderen. She discriminated four types of instruction in collaboration with Amos van Gelderen and Folkert Kuiken. Aartje van Dijk analyzed instruction in the collected studies with feedback from Amos van Gelderen. She wrote the article with feedback from Amos van Gelderen and Folkert Kuiken.

Chapter 3 Writing-to-learn in biology and mathematics teacher education: Promoting students' topic knowledge and insight
This chapter is a slightly modified version of Van Dijk, A., Van Gelderen, A., & Kuiken, F., (2023). Writing-to-learn in biology and mathematics teacher education. Promoting students' topic knowledge and insight. *Frontiers in Education*, *8*. |https://doi.org/10.3389/feduc.2023.1094156

Aartje van Dijk prepared the two quasi experiments in the study in collaboration with two colleague teacher educators from different disciplines. The study was designed and conducted by Aartje van Dijk with feedback from Amos van Gelderen and Folkert Kuiken. The analysis of the collected data took place in collaboration with Amos van Gelderen.
The two think aloud studies were prepared by Aartje van Dijk in collaboration with Amos van Gelderen. Aartje van Dijk executed the think aloud studies. Additionally, she analyzed the results with feedback from Amos van Gelderen. The manuscript was written by Aartje van Dijk with feedback from Amos van Gelderen and Folkert Kuiken.

Chapter 4 Effects of instruction in writing-to-learn on low achieving adolescents in biology and mathematics classes.
This chapter is a slightly modified version of Van Dijk, A., Van Gelderen, A., & Kuiken, F., (2022). Effects of instruction in writing-to-learn on low achieving adolescents in biology and mathematics classes. *Education Sciences*, *12*. https://doi.org/10.3390/educsci12090595

Aartje van Dijk prepared two quasi experiments with feedback from Amos van Gelderen and Folkert Kuiken. She collaborated with four student teachers, interns, who taught grade 7 and grade 10 students in prevocational education. The study was designed by Aartje van Dijk, Amos van Gelderen and Folkert Kuiken.
Two think aloud studies were prepared by Aartje van Dijk with feedback from Amos van Gelderen. Aartje van Dijk executed the studies.

The resulting article was written by Aartje van Dijk with feedback from Amos van Gelderen and Folkert Kuiken.

Chapter 5 General discussion
Aartje van Dijk wrote the general discussion with feedback from Amos van Gelderen and Folkert Kuiken.

The Summary was written by Aartje van Dijk with feedback from Amos van Gelderen and Folkert Kuiken.

De Nederlandse samenvatting (Dutch summary) was written by Aartje van Dijk with feedback from Amos van Gelderen and Folkert Kuiken.

Dankwoord (acknowledgements)

In de eerste plaats bedank ik de Hogeschool Rotterdam voor het faciliteren van de uitvoering van mijn promotieonderzoek. Ook veel dank voor het Kenniscentrum Talent Ontwikkeling dat mijn thuisbasis was gedurende de uitvoering.

Amos van Gelderen, co-promotor, stimuleerde mij om een promotieonderzoek te doen op het moment dat ik op zoek was naar een onderwerp waar ik me in kon verdiepen. Hij stelde schrijvend leren voor. Amos vormde samen met mijn promotor Folkert Kuiken een goed team. Beiden wil ik bedanken voor de begeleiding bij de uitvoering van de studies en het schrijven van het proefschrift. Jullie hebben me door het onderzoeksproces geloodst en me gesteund gedurende een lange periode! In de laatste maanden waren jullie zelfs tweewekelijks present op het beeldscherm om versies van de tekst van het proefschrift te bespreken. *Teams* bleek toen heel nuttig te zijn. Heel hartelijk bedankt!

Veel dank voor mijn collega-lerarenopleiders Damian McCrossan (biologie) en Lennart de Jonge (wiskunde) voor hun bereidheid om schrijvend leren in hun lessen te gebruiken. Hun studenten Sophia de Jong, Bastiaan Nolet (biologie), Deveney Stolk en Isabel van Dijke (wiskunde) waardeer ik ook zeer voor hun enthousiasme om met schrijvend leren aan de slag te gaan in hun eigen (stage) lessen. Verder bedank ik de twee lerarenopleiders van de Hogeschool van Amsterdam, Patricia Kruit (biologie) en Marieke Collins (wiskunde), die vroege versies van schrijfopdrachten hebben uitgeprobeerd in hun lessen.

Enkele studenten Nederlands en een medewerker van Bureau Uitgetypt hebben protocollen van de gefilmde hardopdenksessies uitgewerkt. Met vriendin en oud-collega Ria van Zijl heb ik als vanouds samengewerkt en de protocollen gecodeerd. Banu Alkan, student Nederlands, heeft geholpen bij het invoeren van de data. Zeer bedankt voor jullie bijdragen!

En dan was er nog het thuisfront. Het promotietraject duurde lang en vroeg veel van het geduld van man en kinderen. Tot overmaat van ramp werd ik aangereden, wat tot gevolg had dat Dick, mijn man, veel klussen moest klaren. En dat juist toen we in de voorbereiding van onze verhuizing zaten. Lieve Dick, dank voor je steun en je geduld!

General introduction

1.1 Writing-to-learn Chapter 1

The studies in this dissertation investigate the effects of education using writing-to-learn in different disciplines. The motivation for this research is the observation that many students in different types of education have trouble acquiring insight and knowledge as demanded by the disciplinary programs (Sampson & Phelps Walker, 2012; Hunter & Tse, 2013; Finkenstaedt-Quinn et al., 2017). Apparently, usual ways of learning, such as listening to teachers' explanations or studying textbooks, do not always lead to the acquisition of new insights and topic knowledge about the disciplinary lesson contents. Another way of learning, not well known in education, is writing-to-learn. This use of writing is intended to support students in mastering disciplinary contents presented in education. Writing-to-learn is defined as writing aimed at acquiring insight in and topic knowledge of disciplinary contents (e.g. in geography, history, biology, chemistry or mathematics). Students record their ideas in text, which provides them with the opportunity to read and reread their text, receive feedback (from classmates or teachers) and reflect on the clarity of the formulation of their ideas. This process may result in new insights into disciplinary contents and topic knowledge, so-called knowledge transformation (Bereiter & Scardamalia, 1987).

Research into effects of writing-to-learn is conducted since the eighties. Initially, writing in itself (without specific instruction) was thought to promote learning. This is called the 'strong text theory' (see chapter 2). Results, however, were mixed (Klein, 1999). Bangert-Drowns et al. (2004) suggested that explicit instruction in the use of cognitive and metacognitive strategies in writing lead to positive effects on learning. An example of a cognitive strategy is organizing ideas for planning a text; an example of a metacognitive strategy is reflecting on what one understands and what one has not understood yet (Nückles et al., 2009). Subsequently, many studies on instruction in writing-to-learn were conducted. Reviews of these studies show that various instructions in cognitive or metacognitive strategies were used and often led to positive effects on insight and topic knowledge (Miller et al., 2018; Graham et al., 2020; Hand et al., 2021). These reviews could, however, not determine which elements of instruction were most successful.

The general aim of the present dissertation is to gain further insight into the effects of instruction in writing-to-learn in different disciplines and types of education. This was tested by means of nine different studies. The first is a review of studies, using four theoretically defined types of instruction in writing to stimulate learning in the regular classroom.The studies were conducted in various disciplines (e.g., biology or history) and types of education. Genre Writing, one of the four types of instruction distinguished, appeared to be the most effective in the review.

In continuation of the review, four quasi experimental studies were conducted to investigate effects of Genre Writing. More specifically, Genre Writing

added with instruction in planning and revision seemed a promising approach. Two of the quasi experimental studies took place in biology and mathematics teacher education. The remaining two were conducted in grade 7 biology class and grade 10 mathematics class. Each quasi experiment was accompanied by a think aloud study to investigate the cognitive processes of writing-to-learn underlying students' performance of a writing task. In the following sections the nine studies are described in more detail.

1.2 Review study

The review study was executed in order to acquire an overview of the literature on instruction in writing-to-learn.The results of the review served as a starting point for conducting the four quasi experiments and four think aloud studies.

In the review it was investigated which theoretically defined types of instruction in writing lead to effects on learning. Therefore, experiments, quasi experiments and case studies directed at the effects of instruction in writing-to-learn from grade 5 to higher education were selected. This limitation was based on evidence suggesting that from grade 5, students are able to reflect on their writing (Van Gelderen, 1997). A number of 43 studies was selected. On the basis of theoretical assumptions about underlying (meta)cognitive processes leading to learning (Klein, 1999), three types of instruction were distinguished (Forward Search, Backward Search and Genre Writing). A fourth type of instruction (Planning Only) emerged from the studies found and was added to the systematic analysis.

Forward Search instructs students to first write a draft without planning and thereafter review and revise their draft. Forward Search implies that writers first externalize their pre-existing knowledge into text, as a preparation for reviewing and revising activities. By reviewing and revising directed at rhetorical and conceptual issues, writers can discover new conceptual relations and acquire new insights.

Backward Search instructs students to first make a plan, write a draft and finally to review and revise their draft. During reviewing, students look back to their plan and ask themselves whether their draft aligns with the planned rhetorical goals. If not, they revise their draft or adjust the plan. Such backward search may take place various times until the writer is satisfied. The process of adjusting drafts and plans to each other can change writers' perspective on contents and rhetorical demands of the writing task and therefore lead to new insights and topic knowledge.

Genre Writing instruction (in some reviewed studies added with planning or with planning and revision) requires students to conform to the demands of a specific (text) genre, for instance 'explanation', and to use conceptual relations characteristic for that genre, for instance 'comparing'. For writing in a genre, students need genre knowledge. Therefore, before students start to write, the teacher demonstrates the genre by means of a model text, and points at linguistic realisations of typical conceptual relations. When writing, students need to reflect on how to use these conceptual relations. Students' reflection on genre-specific ways of ordering their ideas and relating them can make them aware of new ways of organizing and

formulating the contents of their drafts and texts, and therefore may result in new insights into the topic of their writing.

Finally, Planning Only instructs students on planning their texts. Planning activities entail selecting and organizing contents, and setting a rhetorical goal. Wallace et al. (2007) consider planning activities a way to arrive at new insights. These activities may be exchanging ideas with other students and making notes to organize contents. By carrying out these types of activities, students may reconsider their selection of content elements for their text to be written. When adapting and reconsidering their planning, new insights and topic knowledge may arise.

The number of experimental comparisons leading to positive effects on measures of topic knowledge and insight were computed for each of the four types of instruction to determine for which type of instruction the most positive evidence was found. This appeared to be Genre Writing, specifically the type added with instruction in planning and revision.

1.3 Four quasi experimental studies

As in the review study the most promising results were found for Genre Writing added with Planning and Revision (GWPR), we decided to adopt GWPR instruction for our further investigations. It stimulates the process of writing-to-learn not only by instruction of how to write in a genre, but also by instruction directed at planning and revising activities. A posttest-only design was used for each of the four studies. The intention was to investigate whether GWPR instruction could be applied in various types of education. Therefore, two studies were conducted in teacher education. The other two studies took place in prevocational education, one in grade 7 classes and one in grade 10 classes.

Additionally, the studies in teacher education and in prevocational education were conducted in two disciplines to investigate effects of GWPR instruction in various disciplines. The two disciplines chosen, biology and mathematics, are different in many ways, but most importantly in the role attributed to writing. In biology, writing is regularly used, for instance for composing lab reports, whereas in mathematics writing is used scarcely. Our expectation was that if GWPR instruction leads to positive effects on learning in such differing diciplines, there is a fair chance that it promotes learning in other disciplines as well.

In each of the four quasi experiments, the writing-to-learn tasks used were co-created in collaboration with the teacher(s). By co-creating with teachers, the writing-to-learn tasks could be well aligned with the lesson objectives. The teachers in prevocational education were in their final internship. Although less experienced than the teacher educators, they were familiar with writing-to-learn, because in the year before the quasi experiment, they had participated in the experimental groups of the studies in teacher education. This facilitated the composition of writing-to-learn tasks that were well adjusted to the courses at stake and overall implementation fidelity.

In the four studies, the experimental conditions were compared to business as usual control conditions. Significant differences in learning in favor of an

experimental group were explained as resulting from GWPR instruction on writing-to-learn.

1.4 Four think aloud studies

It is assumed that cognitive processes underlie students' activities while writing to learn (Bereiter & Scardamalia, 1987). Think aloud studies were directed at identifying indicators of these processes. Therefore, small samples were drawn from the experimental and control students participating in the quasi experiments. In individual, video recorded sessions, each student was asked to think aloud during the performance of the post-test writing task. Students' speech was transcribed and coded using a coding scheme derived from Hayes and Flower (1980). Codes for reflection on planning, formulating or revising texts while writing were considered to indicate the process of writing-to-learn. The numbers of indicators used by the experimental students were compared to those used by the control students. This comparison was made to test the hypothesis that the experimental students manifested more signs of the process of writing-to-learn as a result of their experience with writing-to-learn tasks in the experimental lessons.

1.5 Outline of this dissertation

Because the chapters 2, 3 and 4 are written as separate journal articles, there is considerable overlap in the methods sections of the chapters 3 and 4 and in the theoretical overviews in the chapters 2, 3 and 4. Although that may seem quite redundant to the readers of this dissertation, the advantage is that each of these chapters can be read as a stand-alone text.

Chapter 2 reports the findings of the empirical review of studies on instruction in writing-to-learn, aimed at determining whether the four types of instruction (Forward Search, Backward Search, Genre Writing and Planning Only) lead to effects on disciplinary topic knowledge and insight. Chapter 3 comprises two quasi experiments in teacher education, one in biology and one in mathematics teacher education. Both investigate effects of Genre Writing instruction added with Planning and Revision (GWPR). Each quasi experiment is followed by a think aloud study, directed at identifying indicators of the process of writing-to-learn. The findings of the quasi experiments and the think aloud studies are reported, as well as conclusions, discussion, suggestions for future research and educational implications. Chapter 4 focuses on writing-to-learn by low achieving students in pre-vocational education learning biology and mathematics. Similar to chapter 3, each of the two quasi experimental studies is followed by a think aloud study investigating low achievers' writing-to-learn processes. Findings of the two quasi experiments and think aloud studies are reported, followed by conclusions, discussion and suggestions for further research and educational implications. Chapter 5 provides an overview and a discussion of the findings. Finally, suggestions for future research, limitations and educational implications are discussed

.

Chapter 2

Which types of instruction in writing-to-learn lead to insight and topic knowledge in different disciplines? A review of empirical studies

Chapter 2

Which types of instruction in writing-to-learn lead to insight and topic knowledge in different disciplines? A review of empirical studies[1]

Abstract

This review examines which types of instruction in writing-to-learn lead to effects on insight and topic knowledge in different disciplines, in grades 5 to 12 and in higher education. Forty three empirical studies have been selected to answer this question. Four types of instruction are distinguished. Three of them are based on hypotheses proposed by Klein (1999) about the cognitive processes involved in writing-to-learn: Forward Search, Genre Writing and Backward Search. The fourth type, Planning Only, arises from the literature reviewed. Results of the studies show that about two thirds of the (quasi) experimental studies lead to positive effects on insight and topic knowledge for the four types of instruction. However, given the small number of experimental studies conducted, no firm conclusions can be drawn for three types of instruction. For the remaining type of instruction, Genre Writing, a larger number of studies provides positive evidence. Suggestions for future research are discussed.

[1] This chapter is a slightly modified version of the published article:
Van Dijk, A., Van Gelderen, A., & Kuiken, F., (2022). Which types of instruction in wrting-to-learn lead to insight and topic knowledge in different disciplines? A review of empirical studies. *Review of Education*. https://doi.org/10.1002/rev3.3359

2.1 Introduction

Language is an important mediator of teaching and learning in all disciplines. Teachers explain subject matter by talking, and students study subject matter by listening, discussing and reading. Students' writing is normally used for assessing insight and topic knowledge, but much less for learning (Linnemann, & Stephany, 2014). Writing assignments are generally not used by teachers as a tool for developing insight into subject matter or for acquiring topic knowledge. Baker et al. (2008) described problems teachers foresee for including writing-to-learn in their teaching, such as the extra time that is required and their insecurity in evaluating students' texts (see also Wallace et al., 2007; Bean, 2011). On the other hand, cognitive activities involved in writing might be an important additional option for arriving at deeper insight and new topic knowledge in the disciplines.

Acquiring insight and topic knowledge as the main goals of writing is what we call writing-to-learn, which is the topic of this study. Boscolo and Carotti (2003) describe writing-to-learn as follows: "all the writing activities aimed at facilitating and/or strengthening recall, understanding or elaboration of concepts and ideas" (p. 200). Examples are writing a learning-journal to reflect on subject matter, writing to explain a subject adapted to the needs of readers that have no prior knowledge, or writing an argument based on several texts.

Emig (1977) stated that next to reading, listening and speaking, writing can be an important tool for learning as well. She considers writing a way of encouraging (critical) thinking and understanding, because it requires the writer to discern conceptual relationships, and to display these by means of syntactical, lexical and rhetorical devices. According to Emig (1977), writing provides unique opportunities for learning, because the slow pace of the writing process gives extra room for reflection. Furthermore, the result of the writing process remains visible, enabling feedback on text contents and thereby providing new insights.

In research on writing-to-learn, learning concerns higher level thinking processes, which are described as "integrating new information and prior knowledge" (Rivard & Straw, 2000, p. 567) or as knowledge transforming (Wallace et al., 2007, p. 27, p. 31). The latter is derived from Bereiter and Scardamalia's (1987) writing model that influenced theory on writing-to-learn to date (Klein, & Boscolo, 2016). Applebee (1984) states that writing tasks aiming at writing-to-learn should be focused on "heuristic activity in which subject knowledge is examined and extended" (p. 589).

Apparently, writing-to-learn is considered a comprehensive process, directed at insight into the meaning of newly learned knowledge and recall of topic knowledge. In the present review, we take the view that both types of learning by writing are relevant. By topic knowledge we mean recall of learned concepts (retrieving concepts from memory). By insight we mean the ability to relate new concepts to prior (topic) knowledge. We view insight as the ultimate aim of writing-to-learn, because insight is the manifestation of higher order understanding of concepts.

The present review aims at distinguishing types of instruction in writing that may lead to new insight and topic knowledge. The next two sections respectively discuss early studies into the character of writing-to-learn, and more recent theories about it. Following on from that, the types of instruction are described.

2.2 Theoretical overview

2.2.1 Early studies into writing-to-learn

Since the 1970s, researchers were intrigued by the idea that writing evokes a way of thinking and learning, which can be applied in education. The prevailing view was that any writing task leads to learning, which is called the strong text theory. Writing was seen as "a mode of learning" as Emig (1977) called it in the title of her essay. According to Klein and Boscolo (2016), the focus was on two types of writing: analytical writing (arguments, essays) and personal writing (e.g., diaries, poetry, stories and learning journals). Both types were considered relevant for learning in all disciplines and therefore were summarized in the phrase Writing Across the Curriculum.

Applebee (1984) noticed that the relation between writing and learning was an unexamined assumption. He argued that writing-to-learn might be a more complex process than was thought, and that the specific design of writing assignments might determine what type of learning occurs. He observed that the effect of writing might be retention, for instance when students write answers to questions about newly learned topics, or a new insight, for instance when students write an argument to explain their point of view. Consequently, Applebee (1984) proposed experimental research into the interaction between writing tasks and the aims of teachers. Langer and Applebee (1987) elaborated on this interaction and examined which pedagogical conditions contribute to learning. They concluded that treating content in various ways when writing, for instance when students are required to revise their writing, might result in productive learning about that content in terms of topic knowledge as well as insight.

Early studies into effects of writing-to-learn often are based on the strong text theory, (Ackerman, 1993; Rivard, 1994). These studies that were conducted in diverse grades and disciplines (Ackerman, 1993), were seldom situated in regular classrooms, and in most cases directed at science (Rivard, 1994). At the most, some studies (e.g., Langer, & Applebee, 1987; Boyles et al., 1994) showed a complex picture of positive and negative results. This caused Ackerman (1993) as well as Rivard (1994) to conclude that the strong text theory was not supported by the outcomes of research.

In Bangert-Drowns' et al. (2004) meta-analysis on effects of writing-to-learn, one of the research questions is whether particular types of writing assignments, for instance personal expressive writing, result in learning, thus elaborating on Applebee's view. The authors did not find an effect on learning and concluded that the strong text theory does not hold, in accordance with Ackerman (1993) and Rivard (1994). In addition, Bangert-Drowns et al. (2004) assembled studies that were performed in school settings. They examined 48 studies directed at the contexts of primary education, secondary education and university, 34 of which were conducted in regular classrooms, and in different disciplines. They found some support for their hypothesis that writing assignments including prompts on metacognition, such as reflection on students' own learning processes, have an effect on learning, and might stimulate the process of writing-to-learn. The authors suggested that effects might be

larger when instruction also comprises training of cognitive writing processes before writing, for instance goal setting and organizing.

2.2.2 Theories about the process of writing-to-learn

Three theories about the process of writing-to-learn have been proposed: the knowledge-transforming model by Bereiter and Scardamalia (1987), the dual process model by Galbraith (1992, 2009) and four hypotheses concerning the cognitive processes of writing-to-learn by Klein (1999).

According to the knowledge-transforming model (Bereiter & Scardamalia, 1987), interaction takes place between a content space containing writers' topic knowledge, and a rhetorical space containing their discourse knowledge. In this interaction, writers set rhetorical goals, generate content, revise their rhetorical goals, and repeat these actions, until they consider their text satisfactory. For instance, when writing a text for an audience, a writer describes a theory (content) in abstract terms. When rereading the text, the writer realizes that the audience may not understand it, because extra knowledge about the topic is needed (rhetorical goal). Therefore, the writer decides to give an example (content) for making the text more accessible to the audience. As a result of these interactions between content space and rhetorical space, writers may acquire new insights.

Galbraith (1992) distinguishes two composing styles. He regards writing including planning activities as a composing style of high self-monitors opposed to the approach of writing without planning, which he considers a style of low self-monitors. High self-monitors pay much attention to rhetorical aspects when planning and reviewing, whereas low self-monitors write spontaneously and only attend to rhetorical aspects after reviewing the content of their texts.

Galbraith (2009) tests his ideas empirically, arriving at the dual process model. This model distinguishes a knowledge retrieval system and a knowledge constituting system in writing. Writers use their knowledge retrieval system, in which explicit knowledge is stored, for retrieving and planning content. When writing, they use their knowledge constituting system in order to make new connections between concepts, of which the writer was not aware previously (implicit knowledge). According to Galbraith (1992, 2009), a characteristic of high self-monitors is that they are inclined to hold to their planning while writing, whereas low self-monitors are inclined to deviate from their planning whenever they want. In the latter situation new insights may arise, whereas high self-monitors may produce well-structured texts, but no new insights.

Klein (1999) derives four hypotheses from the literature for how cognitive and metacognitive processes may contribute to writing-to-learn. The first, known as 'writing at the point of utterance' (based on Britton, 1982), states that a text resulting from spontaneous writing reveals writer's knowledge. The writing-to-learn process is supposed to take place during formulating. This hypothesis corresponds to the strong text theory that was rejected earlier by Ackerman (1993) and Rivard (1994) (see previous section). Therefore, from here on we refer to Klein's other three hypotheses. The second hypothesis, Forward Search, is based on Galbraith (1992), among others, and assumes that writers write down their ideas in a first draft without any preparation and keep on writing until they have written all they can think of. Then, they reread their drafts, draw new inferences or recognize a flaw in the logic, and revise them

taking rhetorical goals into account. The assumption is that the process of writing-to-learn takes place in reviewing and revising.

The focus of the third hypothesis, Genre Writing, is on the genre of the text to be written. A genre can be characterized by its rhetorical goals and the relations between text elements directed at attaining these goals (Halliday, & Martin, 1993). For instance, the rhetorical goal of the genre argument is to convince the audience, and the relations between the text elements (opinion, arguments and conclusion) are argumentative. Klein (1999) based the Genre Writing hypothesis on the view that writers composing a text of a specific genre have to use their knowledge of that genre. Depending on the genre, writers formulate their rhetorical goals, and the relations between the elements belonging to the genre. By using their genre knowledge, they may recognize relations between concepts they were not aware of before (Newell, 1984). Thus, by paying attention to (reflecting on) the rhetorical requirements of the genre, writers may acquire new insight into conceptual relations between elements of their topic knowledge (Langer, & Applebee, 1987).

Finally, the fourth hypothesis, Backward Search, implies that writers set rhetorical goals, generate content based on the rhetorical goals (planning), subsequently write their text, and finally revise their text referring to their rhetorical goals and planned content. Just as for the Forward Search hypothesis, the process of writing-to-learn is assumed to take place in reviewing and revising the content. The Backward Search hypothesis however assumes that writers specifically reflect on goals (planning) set before writing. While performing these activities, writers may discover relations between concepts they were not aware of yet, leading to new insights. This - recursory - process is derived from the knowledge-transforming model of Bereiter and Scardamalia (1987).

2.2.3 Types of instruction in writing-to-learn

Since Bangert-Drowns et al. (2004) considered instruction in cognitive and metacognitive writing processes necessary for eliciting the process of writing-to-learn, the following generation of studies explored various types of instruction. Recently, these were reviewed by Miller et al. (2018), Graham et al. (2020) and Hand et al. (2021).

The systematic review of 43 studies by Miller et al. (2018) aimed at exploring the state of research on writing-to-learn in grades 6 to 12. The reviewed studies took place in the disciplines humanities, social studies, science and mathematics. The researchers performed an inductive analysis revealing that instructing writing-to-learn by means of a checklist for organizing and generating activities (Science Writing Heuristic) or by journaling led to positive effects. In addition, they found positive effects of inquiry-based instruction of cognitive and metacognitive writing processes. The positive results were found in 46 percent of the reviewed studies.

In their meta-analysis, Graham et al. (2020), found that 82 percent of 56 reviewed studies in grades 1 to 12 in science, social studies and mathematics led to positive effects on learning with an average effect size of .30. However, the researchers found a large variability in effect sizes ranging from 1.67 to -0.74, which they could not explain by means of a moderation analysis of any of a large number of variables, such as type and features of writing activities and instruction of cognitive

and metacognitive writing processes. Therefore, it was not clear which components of instruction led to the found positive effects.

Hand et al. (2021) reviewed 81 (master and doctoral) theses on the application of instruction using the Science Writing Heuristic. The researchers found that its use resulted in growth of insight and topic knowledge regardless of grade or cultural background. In particular, the reviewers investigated students' knowledge-generating activities for identifying patterns related to the outcomes. Their qualitative analysis showed that the duration of the intervention was a determining factor in arriving at positive effects for students as well as for teachers. Students needed time for mastering knowledge generating and teachers needed time for exploring how to coach students at it. Other influential factors were teacher's critical questions stimulating students' thinking and the combination of individual, group and class activities for performing generating activities.

Thus, the reviews appear to provide evidence for positive effects of instruction in studies on writing-to-learn (Graham et al., 2020), of specific ways of instructing writing-to-learn (Miller et al., 2018, Hand et al., 2021), and of conditions contributing to these outcomes (Hand et al., 2021). The three reviews do not provide insight into the relations between the architecture of instruction and underlying writing-to-learn processes. Graham et al. (2020) investigated various modes of instruction, but is not able to explain the found effects. Miller et al. (2018) as well as Hand et al. (2021) show positive effects of instruction by means of the Science Writing Heuristic, but do not explain their outcomes. The three reviews used a data driven approach.

In the present study, we attempt a more theory-guided approach by systematically discriminating types of instruction on the basis of theoretical assumptions about the (meta)cognitive processes that they invoke. Klein's (1999) three hypotheses served as the basis for describing types of instruction. Klein (1999) composed his hypotheses about the process of writing-to-learn by means of (combinations of) the (meta)cognitive processes planning and reviewing and of the use of genre knowledge during writing. By analyzing examples of concrete instruction, it is possible to discern on which of these processes the approach is based.

Instruction directed at spontaneous writing followed by inspection, feedback and/or revision can be classified as instruction based on the Forward Search hypothesis. In contrast, instruction that emphasizes planning activities before writing, followed by prompts to revise the drafts to ensure a better coverage of the subject or better reception by the readership (instigated by feedback by peers or teachers), can be classified as instruction based on the Backward Search hypothesis. Finally, instruction emphasizing that students are familiarized with genre characteristics of the texts they are supposed to write and that they use these characteristics while writing, can be classified as based on the Genre Writing hypothesis.

This review approach allows investigating the relative effectiveness of each of these three types of instruction. In addition, it allows us to see to what degree the three hypotheses of Klein (1999) cover the types of instruction that have been studied until now. In case that instructional approaches are used that clearly deviate from the three types, this is valuable information that may contribute to a better understanding of underlying assumptions in instructional types directed at writing-to-learn.

2.2.4 The present study

Additional to the outcomes of the previous meta-studies, the present study is aimed at clarifying which cognitive (organizing, generating, goal setting, using genre knowledge) and metacognitive writing processes (reviewing, revising) are part of effective instruction in writing-to-learn. Therefore, we conducted an empirical review. Starting from Klein's (1999) hypotheses, we analyzed studies using instruction based on one of the three types of instruction discerned and their effects on insight and topic knowledge. In addition, we analyzed studies that deviated from the three types and compared them to the other types. Presumably this provides us with better understanding of the cognitive and metacognitive writing processes involved in writing-to-learn and in the types of instruction that can set them off. Our research question is as follows:

> Which types of instruction in writing-to-learn directed at cognitive and metacognitive writing processes result in new insights and topic knowledge in several disciplines?

Because instruction in writing-to-learn is meant to be effective in the context of education, we limited our review to studies taking place in regular classrooms. We were interested in how instruction on writing-to-learn can be embedded in educational practice while improving students' insight and topic knowledge.

2.3 Method

2.3.1 Criteria for selection

Our main objective was to investigate whether instruction directed at different (meta) cognitive writing processes, such as specified by Klein's (1999) hypotheses, was effective for acquiring new insight and topic knowledge in educational contexts. Therefore, our criteria for inclusion focus on aspects of instruction and education. Regarding the design of the studies, we preferred a liberal policy and included experimental studies as well as case studies. In doing so, we included empirical studies which applied as many various modes of instruction as possible, in as many types of educational contexts as possible. More specifically, we used the following criteria for including studies:

- The study is an empirical study (experimental, quasi experimental, or a case study).
- The study is aimed at writing-to-learn.
- The study measures effects on insight and/or topic knowledge concerning a disciplinary subject.
- The study is embedded in a regular classroom context.
- The study is directed at grade 5 or higher.
  Grade 5 students having acquired the basic principles of writing are able to reflect on their own and other students' writing and to revise texts when instructed and stimulated (Van Gelderen, 1997). This reflection can be considered conditional for the processes of writing-to-learn (Bereiter & Scardamalia, 1987).

- Instruction in writing-to-learn is clearly described (writing assignments explaining students what to write about and instruction how to proceed when writing).
- The pre- and post-tests are clearly described.
  Because we wanted to investigate effects of instruction on insight and topic knowledge, we needed to know what exactly was measured and how. For tests comprising open-ended questions or multiple-choice items, the aimed outcomes are respectively considered insight and topic knowledge, unless researchers explicitly aimed otherwise (e.g., multiple choice items measuring insight). For other measures, for instance writing assignments, assessment criteria should be directed at insight (e.g., students can provide a summary of lesson contents) and/or topic knowledge (e.g., students can recall lesson contents).

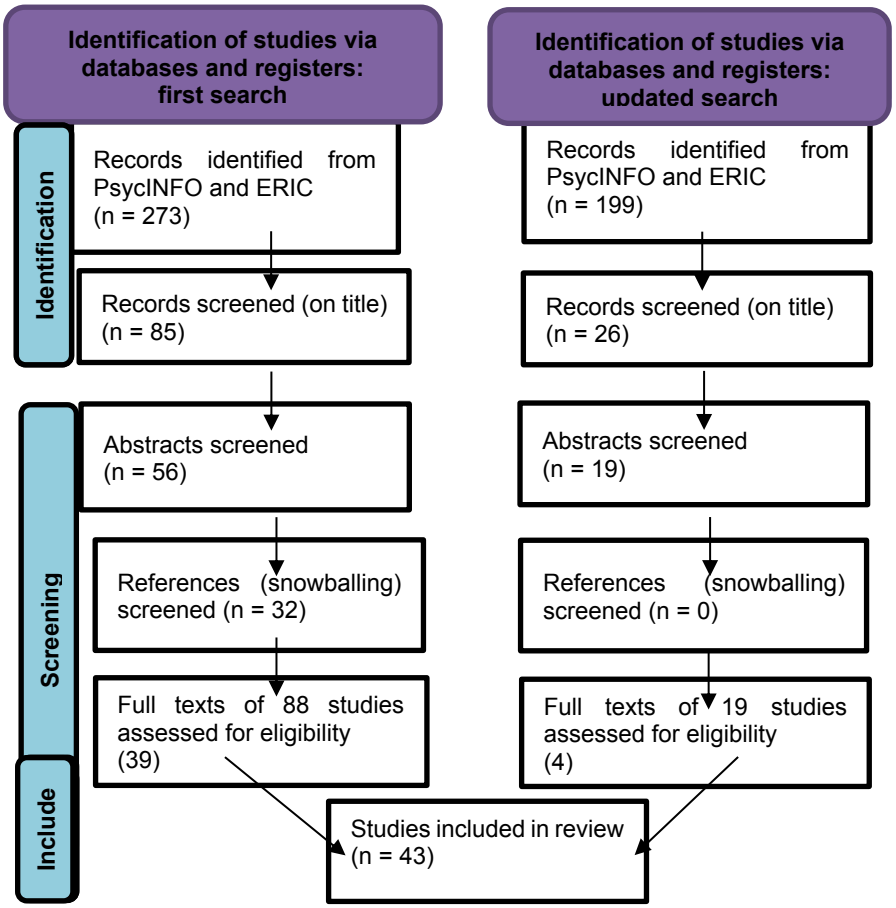
The following criterion was used for exclusion:

- The study is directed at gifted students only or at special needs students only.

2.3.2 Procedure for search and selection of literature

We used automatic and manual search methods. Two searches were undertaken. The starting point of the first search was 1999, the year of publication of Klein's review indicating a reorientation in research on writing-to-learn, resulting in an important role for instruction to evoke learning by writing. Combined in one search, the databases ERIC and PsycINFO were consulted for published studies. In order to ascertain that no relevant studies were missed, we systematically searched the bibliographies of the studies that were found relevant for additional studies (so-called snowballing). This search ended in September 2018. The second search provided an update and covers the period from September 2018 to August 2021.The databases ERIC and PsycINFO were consulted for published studies. Additionally, we systematically searched the bibliographies of the studies that were found relevant (snowballing). For an overview of the searches, see Figure 1 (based on Page et al., 2020).

Figure 1 Flow diagram of the searches



To secure the validity of the selection procedure (see below), the three authors discussed the selection of studies proposed by the first author and decided on the basis of consensus whether studies should be included and for which of the discriminated types of instruction they were representative. An example is Nevid et al. (2017) using elaborated writing assignments. The question was whether the writing assignments also included some type of instruction directed at writing-to-learn. The conclusion was that this was not the case. The writing assignments described the context of the task precisely but did not provide (meta)cognitive procedures for the writing process (such as planning, revising or using genre knowledge). Therefore, the study was not included. Another example is Mateos et al. (2018), a study directed at learning to write syntheses. Students were required to write about contents from sources some of which were contradicting each other. The question was whether the assignment was directed only at learning to write or comprised writing-to-learn as well. It was decided that the latter was the case, because next to writing about contents from sources, students also

had to reflect on how to reconcile conflicting contents when writing their synthesis. Therefore, this study was included.

In both searches, we used the following descriptors: writing-to-learn, learning by writing, writing as a learning… (the latter is the start of a phrase), insight writing, deep learning writing, critical thinking writing, writing in the disciplines, disciplinary writing, written argumentation, argument writing, topic knowledge writing and synthesis writing. This yielded a total of 273 records, from which a first selection was made based on the titles. This resulted in a selection that consisted of 85 studies. For a further refinement, we consulted all abstracts, which resulted in 56 studies. After snowballing in the references of these studies, we added another 32 leading to a total of 88 studies. Finally, after consulting the 88 full texts, 39 studies satisfied all inclusion criteria.

In August 2021 a new search was undertaken for updating the selection (see Figure 1). From 199 records a selection based on the titles was made, resulting in 26 studies. Next, we consulted all abstracts, which resulted in a selection of 19 studies. After snowballing in the references no studies were added. Finally, after reading the 19 full texts, four studies were added to the selection from the first search, resulting in a total of 43 studies for this review.

2.3 Results

2.3.1 Characteristics of the 43 studies

Table 2.1 provides an overview of all 43 studies that were selected for our review. The studies are ordered alphabetically according to the name of the first author. From Table 2.1, it can be inferred that 23 studies are related to knowledge about science, eight to behavioral sciences, six to humanities, five to social studies, and one to earth sciences.

In the 43 studies, 22 different genres are involved, for instance narrative, laboratory report and argumentative letter. The grade levels of students targeted in the studies vary from below grade 7 (six studies), between grade 7 and 12 (24 studies) to higher education (13 studies). The sample sizes vary from 50 and below (13 studies), between 50 and 100 (18 studies) and above 100 (12). For a few exceptions, in the large majority of selected studies more than five and less than 24 lessons were given. Only in three studies the number of lessons is much larger, namely in two studies it is 40 and in one 56.

The research design of 19 studies is quasi-experimental, and of five studies experimental with randomization on the individual level. These 24 studies use a control group for comparison. Five of these apply teachers' usual program for control (business as usual). In 19 studies, the control group is adapted with regard to time on task and/or the content of the program in order to achieve more experimental control. In ten studies there are different experimental groups that are compared to each other. Five of these use randomization on the individual level. The remaining nine studies are case studies.

In 20 studies, the post-tests are aimed at measuring insight as well as topic knowledge (on a combined measure or two separate measures). The post-tests and the delayed post-tests often measure the effects of writing-to-learn by analyzing students'

texts. In 34 studies, covariates (pre-tests measuring insight and/or topic knowledge) are employed. In three studies, last semesters' grades are added to the pre-test scores or used as the only covariate. In nine studies no covariates are used.

In the last column of Table 2.1 (see next page), effects of the treatments in the studies are presented. In case of comparisons between conditions, only significant effects are shown, using the condition numbers involved (as shown in the column named experimental conditions and treatments). In the case that comparisons involve differences in growth between conditions, this is indicated by the term 'growth'. Effect sizes (Cohen's d, eta squared, or partial eta squared) are reported in the majority of the studies. In the cases that these were not reported, we computed Cohen's d when significant effects were found in comparisons between experimental and control groups or between different experimental groups.

Table 2.1 Overview of 43 studies[1]

Study	Discipline/ Genre(s)	Grade level	N	Period+ lessons	Writing assign-ments	Teacher Researcher or both
Study 1 Atasoy, & Küçük (2020)	Physics Answers on open-ended questions	Grade 8	18	6 weeks 24 lessons	6	Teacher Researcher
Study 2 Balgopal, Wallace, & Dahlberg (2012)	Science Essay	Undergrads	89	NA	3	Teacher
Study 3 Boscolo, & Carotti (2003)	Literature Commen tary	Grade 9	50	28 weeks 56 lessons	12	Teacher
Study 4 Corcelles Seuba, & Castello (2015)	Philosophy Argument	Grade 11	6	8 weeks 7 lessons	3	Teacher
Study 5 Finkenstaedt Quinn, Halim, Chambers, et al. (2017)	Chemistry Summary	Juniors Seniors	36	3 weeks	2	Teacher 2 assis-tants
Study 6 Granado-Peinado, Mateos, Martin, Cuevas (2019)	Psychology Synthesis	Undergrads	160	6 weeks 6 lessons	2	Researcher

1) *For each study, the descriptions of the characteristics are continued in the second part of Table 2.1 on the next page.*

Experimental conditions and treatments	Control group	Randomization	Posttest+ delayed posttest	Covariates or pretest	Significant results
Study 1 Writing about disciplinary topics and students' epistemological views on it.	–	–	1. IN [2] (interview, 5 questions)	IN (interview, 5 questions)	Growth on 1 (5 questions): 30% of the students
Study 2 Writing three texts, each another aim, from viewpoint of one type of person, using a checklist	–	–	1. IN (WT)	IN (WT)	Growth on 1: 33 % of the students
Study 3 Writing and discussing different genres aimed at literary comprehension	Business as usual	–	1. TK (WT) 2. IN (WT)	TK (WT) IN (WT)	Exp. cond > ctr. on 2: partial η^2= .22
Study 4 Collaborative writing using planning guide	–	–	1. IN (WT, individual)	IN (WT, individual)	Growth on 1
Study 5 Writing for a non - expert audience using double-blind peer review, revision of initial draft	–	–	1. IN (MC) 2. IN (WT)	IN (MC) IN (WT)	Growth of exp. cond from initial draft to final draft on 2
Study 6 1. Instruction of checklist for synthesis writing and collaborative learning. Collaborative writing 2. Instruction of checklist for synthesis writing. Collaborative writing 3. Collaborative writing, using checklist 4. Collaborative writing	–	+ (individual)	1. IN (WT, pairs) 2. IN (WT, individual)	IN (WT, pairs) IN (WT, individual)	Cond 1 > cond 3, on 1 and 2, Cohen's *d* = 1.7 Cond 1 > cond 4 on 1 and 2, Cohen's *d* = 2.03 Cond 2 > cond 4 on 1 and 2 Cohen's *d* =.79

[2)] *Abbreviations: cond* condition, *ctr* control group, *exp* experimental, *IN* insight, *MC* multiple choice, *NA* not available, *OE* open ended questions, *quest* questionnaire, *TK* topic knowledge, *WT* writing task, > outperforms, + means present, - means not present, * two or more conditions taken together

Study	Discipline/ Genre(s)	Grade level	N	Period+ lessons	Wri-ting assign-ments	Teacher, researcher or both
Study 7 Gunel, Hand, & Gunduz (2006)	Physics Synthesis	Grade 11	132	2 weeks 10 lessons	2	Teacher
Study 8 Gunel, Hand, & McDermott, (2009)	Biology Explanation	Grade 9, Grade 10	118	10 lessons	1	Researcher
Study 9 Hand, Gunel, & Ulu (2009)	Physics Explanation	Grade 10	181	21 lessons	2	Teacher
Study 10 Hand, Hohenshell, & Prain (2004)	Biology 1. Explana-tion 2. Newspaper article	Grade 10	73	6 weeks 6 lessons	2	Teacher English teacher

For each study, the descriptions of the characteristics are continued in the second part of Table 1 on the next page.

Experimental conditions and treatments	Control group	Randomization	Posttest + delayed posttest	Covariates or pretest	Significant results
Study 7 Writing for teacher: 1. Power p. + script (*sample A*) 2. Plain paper (*sample B*) Writing for grade 10: 3. Power p. + script (*sample A*) 4. Plain paper (*sample B*)	–	+ (group)	1.TK (MC) 2. IN (OE)	1.TK (MC) 2. IN (OE)	Cond 1 > cond 2 on 1: Cohen's *d* = .2, and on 2: Cohen's *d* = .6 Cond 3 > cond 4 on 1: Cohen's *d* = .8, and on 2: Cohen's *d* = .6
Study 8 Writing for: 1. 3rd/4thgraders 2. parents 3. peers 4. teacher	–	+ (group)	1. TK (MC) 2. IN (OE)	1. Last years' biology grade 2. TK, IN (MC, OE) 3. TK, IN (MC, OE)	Growth of cond 1 – 4 on 2: partial η^2 = .08
Study 9 Writing 2 texts using: 1. math, math 2. math, graph 3. graph, graph 4. graph, math	Con - trolled	+ (group)	TK + IN (MC)	1. TK + IN (MC) 2. TK + IN (MC)	Cond 2 > cond 1: *d*= .4 cond 3: *d*= .5 cond 4: *d*= 1.0 ctr: *d*= .5 Cond 1 > cond 3: *d*= .2 cond 4: *d*= .6 ctr: *d*= .2 Ctr > cond 4: *d*= .5
Study 10 1. generating, organizing, then writing, 2 texts 2. generating, writing, then organizing, 2 texts 3. generating, organizing, then writing, 1 text 4. generating, writing, then organizing, 1 text	–	+ (group)	After writing one text: 1. IN (WT) 2. IN (OE) After writing two texts: 3. IN (OE) Delayed: 4. IN (OE)	Last semesters' biology grade	Cond 1 and 3*> cond 2 and 4* on 1: *d*= .49 Cond 1 and 2*> cond 3 and 4* on 3: *d*= .70 on 4: *d*= 1.09

Study	Discipline/ Genre(s)	Grade level	N	Period + lessons	Writing assign-ments	Teacher, researcher or both
Study 11 Hand, Hohenshell, & Prain (2007)	Biology 1. Explanation 2. Newspaper article	Grade 10	87	6 weeks	2	Teacher Researcher
Study 12 Hand, Wallace, & Yang (2004)	Biology Summary	Grade 7	93	8 weeks 40 lessons	1	Teacher
Study 13 Hand, Yang, & Bruxvoort (2007)	Chemistry Business letter	Grade 11	52	2 lessons	1	Teacher
Study 14 Hohenshell, & Hand (2006)	Cell biology Laboratory report	Grade 9, Grade 10	91	7 weeks	7	Teacher

For each study, the descriptions of the characteristics are continued in the second part of Table 2 1 on the next page.

Experimental conditions and treatments	Con-trol group	Rando mizati on	Posttest + delayed posttest	Covariates or pretest	Significant results
Study 11 Writing two texts (a, b) with or without nstruction: 1a. no instruction, then (1b) reviewing by audience 2a. generating, organizing, reviewing by audience, then (2b) reviewing by audience 3a. reviewing by audience, then (3b) generating, orga-nizing, reviewing by audience 4a. no instruction, then (4b) generating, organizing, reviewing by audience	–	–	After writing first text: 1. IN (OE) 2. IN (OE) 3. IN (OE) 4. IN (OE) After writing two texts: 5. IN (OE) 6. IN (OE) 7. IN (OE) 8. IN (OE)	Last school year's grade Before writing first text: 1. IN (OE) 2. IN (OE) 3. IN (OE) 4. IN (OE) Before writing second text: 5. IN (OE) 6. IN (OE) 7. IN (OE) 8. IN (OE)	Cond 1a > 2a on 1: η^2= .162 Cond 2a > 1a on 3: η^2=.222 Cond 3a > 4a on 2: η^2= .296, on 3: η^2= .097 on 4: η^2= .255 Cond 1a > 2a and 4b on 1: η^2= .71 Cond 1b, 3a and 2a+2b*> 4a on 6: η^2= .215, on 7: η^2= .141 on 8: η^2= .183
Study 12 Writing using: 1. checklist 2. checklist, and writing for an audience	Con-trolled	–	1. TK (MC) 2. IN (OE)	1. TK (MC) 2. IN (OE)	Cond 1 > ctr on 1: d= 0.15 Cond 2> ctr on 1: d= 0.29 Cond 2 > ctr on 2: d= 1.02
Study 13 Writing for an audience.	Con-trolled	+ group	1. TK (OE) 2. IN (OE) 3. IN (OE) 4. IN (OE)	1. TK (OE) 2. IN (OE) 3. IN (OE) 4. IN (OE)	Exp. cond > ctr on 4: d=0.83
Study 14 Writing using checklist, for: 1. teacher 2. peers	Con-trolled	–	1. TK (MC) 2. IN (OE)	1. TK (MC) 2. IN (OE)	Cond 1 and 2*> ctr on 2: partial η^2= .114

Study	Discipline/ Genre(s)	Grade level	N	Period+ lessons	Writing assign-ments	Teacher, researcher or both
Study 15 Hunter, & Tse (2013)	Economics Essay	University students	1031	8 weeks 6 lessons	2	Teacher Tutors
Study16 Kabataş, Memiş, & Öz (2017)	Science Text with embedded formulas or graphs	Grade 5	32	NA	2	Teacher
Study 17 Kieft, Rijlaars-dam, & van den Berg (2006)	Dutch literature Literary review	Grade 10	113	5 weeks 5 lessons	5	Teacher
Study 18 Kieft, Rijlaars-dam, & van den Berg (2008)	Dutch literature Literary review	Grade 10	113 (of 220)	5 weeks 5 lessons	5	Teacher
Study 19 Klein, & Ehrhardt (2015)	Science 1 Argument 2 Discussion	Grade 7, Grade 8	72	10 days	1	Teacher Research assistent
Study 20 Klein, & Rose (2010)	Science 1 Argument 2 Explana-tion	Grade 5, Grade 6	34	20 weeks	16	Researcher Teacher

For each study, the descriptions of the characteristics are continued in the second part of Table 1 on the next page.

Experimental conditions and treatments	Control group	Randomization	Posttest+ delayed posttest	Covariates or pretest	Significant results
Study 15 Writing text 1: discussion of content, instruction on genre structu-ring Writing text 2: discussion of teacher's feedback on WT1 (= feed forward)	Ctr 1 Ctr 2 Business as usual	–	After four weeks: 1. IN (WT 1) After eightt weeks: 2. IN (WT 2)	–	Growth of exp. cond. on 2: Exp cond > ctr 1 on 2: d= 1,23 Exp cond > ctr 2 on 2: d=0.07
Study 16 Writing by using multi-modal representations	Controlled	+ (group)	1. IN (OE)	IN (OE)	Exp. cond > ctr on 1: η^2= .449
Study 17 1. First generating and organizing, then writing 2. First generating, then writing, then organizing	–	+ (individual)	1. IN (OE)	1. IN (OE)	Cond 1 > cond 2 on 1: d=0.03
Study 18 1. First generating and organizing, then writing 2. First generating, then writing, then organizing	–	+ (individual)	1. IN (OE)	IN (OE)	–
Study 19 1.Stepwise instruction on elaborating writing subgoals 2. Clustered instruc-tion on writing goals	–	+ (individual)	1. TK (MC)	1. TK (MC)	–
Study 20 Instruction of genre knowledge and cognitive strategies	Controlled	+ (group)	1. TK, IN (OE)	1. TK, IN (OE)	Exp. cond > ctr on 1: partial η^2= .21

.

Study	Discipline/ Genre(s)	Grade level	N	Period+ lessons	Writing assign-ments	Teacher, researcher or both
Study 21 Klein, & Kirkpatrick (2010)	Science 1. Argument 2. Explana-tion	Grade 5, Grade 6	113	24 weeks 20 lessons	16	Teacher
Study 22 Klein, & Samuels (2010)	Science Argument	Grade 7, Grade 8	60	20 weeks 40 essons	20	Teacher
Study 23 Martinez, Mateos, Martin, et al. (2015)	History Synthesis	Grade 6	62	5 weeks 15 lessons	3	Researcher Teacher
Study 24 Mateos, Martín, Cuevas, et al. (2018)	Psychology Synthesis	Under-grads	114	3 lessons	2	Researcher
Study 25 McDermott, & Hand (2013)	Chemistry Text with embedded formulas or graphs	Grade 10, Grade 11, Grade 12	70	NA	2	Teacher

For each study, the descriptions of the characteristics are continued in the second part of Table 1 on the next page.

Experimental conditions and treatments	Control group	Randomization	Posttest + delayed posttest	Covariates or pretest	Significant results
Study 21 Instruction of genre knowledge and cognitive strategies	Controlled	+ (group)	1. TK, IN (Cloze, OE)	1. TK, IN (OE)	–
Study 22 Instruction of genre knowledge	Controlled	+ (group)	1. TK, IN (Cloze, OE)	1. TK, IN (OE)	–
Study 23 Instruction of strategies on synthesis writing, resulting in checklist for writing	Controlled	+ (group)	1. TK, IN (OE)	1. TK, IN (OE)	Exp. cond > ctr on 1: partial η^2= .64
Study 24 1.Instruction of checklist for synthesis writing. Collaborative writing using checklist 2. Self-study of checklist for synthesis writing. Collaborative writing using checklist	–	+ (individual)	1. IN (WT, individual)	2. IN (WT, individual)	Cond 1 > cond 2 on 1: η^2 = .13
Study 25 Instruction of embedding graphs and formulas in text, resulting in checklist for self-assessing the writing	Controlled	+ (group)	1. TK, IN (MC, OE) (after unit 1) 2. TK, IN (MC, OE) (after unit 2)	TK (MC)	Exp. cond > ctr on 2: d= .53

Study	Discipline/ Genre(s)	Grade level	N	Period+ lessons	Writing assign-ments	Teacher, researcher or both
Study 26 McDermott, & Hand (2013)	Chemistry Text with embedded formulas or graphs	Grade 10, Grade 11, Grade 12	95	NA	1	Teacher
Study 27 Nam, Choi, & Hand (2011)	Geology Explanation	Grade 8	345	20 weeks 8 lessons	4	Teacher
Study 28 Nevid, Pastva, & McClelland (2012)	Introductory psychology Reflection	Undergrads	135	20 weeks	16	Researcher
Study 29 Nevid, Pastva, & McClelland (2012)	Introductory psychology Reflection	Undergrads	55	20 weeks	16	Researcher
Study 30 Nückles, Hübner, Dümer, & Renkl (2010)	Psychology Learning journal	Undergrads	50	16 weeks 12 lessons	12	Researcher
Study 31 Nückles, Hübner, Dümer, & Renkl (2010)	Psychology Learning journal	Undergrads	62	16 weeks 12 lessons	12	Researcher teacher
Study 32 Ritchie, Tomas, & Tones (2011)	Biology Stories	Grade 6	55	6 weeks	2	Teacher

For each study, the descriptions of the characteristics are continued in the second part of Table 1 on the next page.

Experimental conditions and treatments	Control group	Randomization	Posttest+ delayed posttest	Covariates or pretest	Significant results
Study 26 Instruction of embedding graphs and formulas in text, resulting in checklist for self-assessing the writing	Controlled	+ (group)	1. TK, IN (MC, OE)	TK (MC)	Exp. cond > ctr on 1: d= .62
Study 27 Writing using checklist: 1. sample A 2. sample B 3. sample C	Business as usual	–	TK, IN (OE)	IN (OE)	Cond 1 > ctr d=. 61 Cond 2 > ctr d= .64
Study 28 Choice of 16 topics out of a long list	–	–	1. TK (MC)	–	Exp. cond: topics students have written about > other topics
Study 29 Obligatory topics Choice between two genres	–	–	1. TK (MC)	–	Exp. cond: topics students have written about > other topics
Study 30 Writing using cognitive and metacognitive hints	Controlled	+ (individual)	1. IN (OE) after 8 weeks 2. IN, (OE) after 16 weeks	–	Cond > ctr on 1: partial η^2= .08
Study 31 Writing using: 1. cognitive and metacognitive hints 2. fading cognitive and metacognitive hints	Controlled	+ (individual)	1. IN (OE) after 8 weeks 2. IN (OE) after 16 weeks	–	Cond 1 and 2 > ctr on 1: partial η^2= .09 Decrease of cond 1 from 1 to 2: partial η2= .15
Study 32 Writing biology narratives	Controlled	–	1.TK, IN (BioQuiz questionnaire)	1.TK, IN (BioQuiz questionnaire)	Exp cond > ctr on 1: d= .39

Study	Discipline/ Genre(s)	Grade level	N	Period+ lessons	Writing assign-ments	Teacher, researcher or both
Study 33 Rivard, & Straw (2000)	Ecology Explanation	Grade 8	22	5 weeks 5 lessons	5	Teacher
Study 34 Rivard (2004)	Ecology Explanation	Grade 8	77	5 lessons	5	Teacher
Study 35 Sampson, Phelps, & Walker (2012)	Chemistry Research report	Undergrads	18	20 weeks 15 lessons	4	Teacher
Study 36 Stewart, Myers, & Culley, (2010)	Psychology Micro-theme	Juniors, Seniors	73	5 weeks 10 lessons	10	Teacher
Study 37 Van Drie, Braaksma, & Van Boxtel (2015)	History Argumenta-tive letter	Grade 11	42	5 lessons	2	Researcher Teacher

For each study, the descriptions of the characteristics are continued in the second part of Table 1 on the next page.

Experimental conditions and treatments	Control group	Randomization	Posttest+ delayed posttest	Covariates or pretest	Significant results
Study 33 Discussion before writing	Controlled	+ (individual)	1. TK (MC, OE, concept maps) 2 IN (MC, OE, concept maps) Delayed: 3. TK (MC, OE, concept maps) 4. IN (MC, OE, concept maps)	1. TK, IN (MC, OE, concept maps).	Exp cond > ctr on 1: d=1.0 and on 3: d=1.21 Exp cond > ctr on 2: d= .60 and on 4: d=1.09
Study 34 Discussion before writing	Controlled	+ (individual)	1.TK (MC) Delayed: 2.TK (MC)	TK (MC)	–
Study 35 Writing 4 initial drafts and 4 revised versions using a checklist and a double-blind peer-review guide	–	–	1. TK + IN (WT: initial drafts of 4 WT's) 2. TK + IN (4 revised versions of 4 WT's)	–	Growth on 2 > growth on 1 for all texts
Study 36 Teacher's feedback and grade after each assignment to: 1. non-honors 2. honors	Business as usual (non - honors only)	–	1. TK (MC) 2. TK, IN (WT)	–	Cond 1 > ctr on 1: d = .89, and on 2: d = .60
Study 37 1.Disciplinary argument writing 2. Non-disciplinary argument writing	–	+ (individual)	1. TK (OE) 2. IN (WT)	1.TK (OE)	Growth of cond 1 and 2 on 1

Study	Discipline/ Genre(s)	Grade level	N	Period+ lessons	Writing assign-ments	Teacher, researcher or both
Study 38 Wäschle, Gebhardt, Oberbusch, & Nückles (2015)	Immunology Learning journal	Grade 7	46	3 weeks 6 lessons	3	Teacher
Study 39 Wäschle, Gebhardt, Oberbusch, & Nückles (2015)	Philosophy Learning journal	Grade 10	24	8 weeks 6 lessons	6	Teacher
Study 40 Winstead, Fry, & Villagomez (2012)	Pedagogics Learning journal	Juniors, Seniors	53	15 weeks 15 lessons	15	Researcher
Study 41 Wissinger, & De La Paz (2016)	History Essay	Grade 6 Grade 7	151	3 weeks 15 lessons	3	Teacher 3 assistants
Study 42 Wong, Kuperis, Jamieson, Keller, & Cull-Hewitt (2002)	English literature Response journal	Grade 12	48	4 weeks 9 lessons	2	Teacher
Study 43 Yildiz & Akdag (2020)	History Story Column	Teacher education	70	6 weeks 6 lessons	6	Resear-chers

For each study, the descriptions of the characteristics are continued in the second part of Table 1 on the next page.

Experimental conditions and treatments	Control group	Randomization	Post-test + delayed post-test	Covariates or pretest	Significant results
Study 38 Writing using cognitive and metacognitive hints.	Controlled	–	1. IN (OE) Delayed: 2. IN (OE) 3. IN (WT)	TK, IN (OE)	Exp. cond > ctr on 1: partial η^2= .18 Exp. cond > ctr on 2: partial η^2= .31 Exp. cond> ctr on 3: partial η^2= .24
Study 39 Writing using: 1. cognitive and metacognitive hints 2. cognitive, metacognitive, hints and motivator	–	–	1.TK, IN OE) 2. IN (WT)	–	Cond 2 > cond 1 on 2: partial η^2 = .41
Study 40 Teacher's feedback on each text after each class.	Controlled	–	TK (MC)	TK (MC)	–
Study 41 Discussing argumentation schemes before writing	Business as usual	+ (individual)	TK (MC)	TK (MC)	Exp. cond > ctr on 1: partial η^2 = .23
Study 42 Writing using: 1. general analytic prompts 2. specific analytic prompts	Controlled	+ (group)	1. IN (OE) 2. IN (OE)	–	Cond 1 > ctr on 1: d=1.30 cond 2 > ctr on 1: d= .84 Cond 1 > ctr on 2: d= .59 cond 2 > ctr on 2: d= 1.15
Study 43 1. instruction on collaboratively writing a story for peers 2. instruction on collaboratively writing a column for peers	–	+ (individual)	1. IN (OE)	IN (OE)	Cond 1 > cond 2: η^2 = .13

2.3.1 Four types of instruction

Table 2.2 shows four types of instruction that were distinguished in the selected studies, Forward Search, Genre Writing, Backward Search and Planning Only. Three of them are based on Klein's hypotheses as described in the introductory section. A large part of the studies can be classified by means of the three hypotheses proposed by Klein (1999). However, we found a substantial number of studies that deviated from these hypotheses, because the instruction was based solely on planning for writing. There was no attempt in these studies to combine planning and revision, such

as is the case in studies classified as Backward Search. Therefore, we added a fourth type of instruction in our review, called Planning Only.

Table 2.2 describes the four types of instruction in the 43 studies. We used the processes planning and reviewing based on the writing model of Hayes and Flower (1980) and consistent with Klein's (1999) hypotheses, to specify the more precise processes that were involved in each study. Planning processes involve organizing, and/or generating. Reviewing processes involve the following four instructed activities 1) feedback by (type of readers that provide feedback), 2) feedback on (the focus of feedback given), 3) revising (whether it occurs or not) and 4) focus of revision (what students attend to while revising).

The first type of instruction is aimed at eliciting the Forward Search process. Students are told to start writing their ideas without planning and to review their draft afterwards. This instruction focuses on reviewing to reinforce thinking about the contents of the draft. The second type of instruction, Genre Writing, aims at genre knowledge relating to the genre that students have to write. Instruction focuses on the genre-specific structure of a model text. Furthermore, instruction may also explain the linguistic elements that need to be used to realize a genre-specific way of writing. The third type of instruction is aimed at Backward Search. This type of instruction aims at planning activities, and at reviewing activities that are focused on the rhetorical and content goals set in the planning activities. The fourth type of instruction in Table 2, concerns instruction on planning activities only.

As Table 2.2 shows, some studies have characteristics of more than one type of instruction as distinguished above. Nevertheless, these studies are attributable to one dominant type based on the authors' descriptions of their main interests.

Table 2.2 Types of instruction[1]

Author, date	Experimental conditions	Planning	Feedback		Reviewing	
			Feedback by	Feedback on	Revising	Focus of revision on
Forward Search instruction						
1 Atasoy & Küçük (2020)	Cond 1:	–	Peer, teacher and class	+	+	Students' Epistemological views
5 Finkenstaedt-Quinn et al. (2017)	Cond 1:	–	Peers	+	+	Comprehen-sibility to audience Discipline-based charac-teristics
8 Gunel, Hand, & McDermott (2009)	Cond 1: Cond 2: Cond 3: Cond 4:	– – – –	3rd/4th grade Parents Peers Teacher	+ + + +	+ + + +	All Conditions. Comprehen-sibility to audience
36 Stewart, Myers, &	Cond 1: non honors	–	Teacher	Both Cond: topic know	–	–

Author, date	Experimental tal conditions	Planning	Feedback		Reviewing	
			Feedback by	Feedback on	Revising	Focus of revision on
Culley, (2010)	Cond2: honors	–	Teacher	ledge, application of course concept	–	–
40 Winstead Fry, & Villagomez (2012)	Cond 1:	–	Teacher	Reflecting	–	–
42 Wong, Kuperis, Jamieson, Keller, & Cull-Hewitt (2002)	Cond 1: general analytic prompts	–	Class and teacher	Students' responses to the prompts	–	–
	Cond 2: specific analytic prompts	–	Class and teacher	Students' responses to the prompts	–	–
Genre Writing instruction						
2 Balgopal, Wallace, & Dahlberg (2012)	Cond 1:	Generating	–	–	–	–
3 Boscolo & Carotti (2003)	Cond 1: Writing the genre literature commentary	–	–	–	–	–
20 Klein, & Rose (2010)	Cond 1:	Generating Organizing	–	–	–	–
21 Klein, & Kirkpatrick (2010)	Cond 1:	Generating	Self	+	+	Genre-based characteristic
22 Klein, & Samuels (2010)	Cond 1: Writing the genre argument	–	–	–	–	–
23 Martinez, Mateos, Martin, & Rijlaarsdam (2015)	Cond 1:	Generating Organizing	Self	–	+	–
25 McDermott & Hand (2013)	Cond 1: 2 writing assignments	–	Self, an audience other than the teacher	+ –	+ +	Embeddedness
26 McDermott, & Hand (2013)	Cond 1: 1 writing task	–	Self, an audience other than the teacher	+ –	+ +	Embeddedness

Author, date	Experimental conditions	Planning	Feedback		Reviewing	
			Feedback by	Feedback on	Revising	Focus of revision on
30 Nückles, Hübner, Dümer, & Renkl (2010)	Cond 1:	Generating	–	–	–	–
31 Nückles, Hübner, Dümer, & Renkl (2010)	Cond 1: (meta) cognitive hints	Generating	–	–	–	–
	Cond 2: same hints faded	Generating	–	–	–	–
32 Ritchie, Tomas, & Tones (2011)	Cond 1:	Generating	Peers	–	+	–
37 Van Drie, Braaksma, & Van Boxtel (2015)	Cond 1: Disciplinary writing	–	–	–	–	–
	Cond 2: Non-disciplinary writing	–	–	–	–	–
38 Wäschle, Gebhardt, Oberbusch, & Nückles (2015)	Cond 1:	Generating	–	–	–	–
39 Wäschle, Gebhardt, Oberbusch, & Nückles (2015)	Cond 1: (meta) cognitive hints	Generating	–	–	–	–
	Cond 2: personal utility prompt	Generating	–	–	–	–
Backward Search instruction						
4 Corcelles Seuba, & Castello (2015)	Cond 1:	Generating Organizing	Peers, teacher	+	+	Text structure, content
6 Granado-Peinado, Mateos, Martin & Cuevas (2019)	Cond 1: Instruction by means of video modeling in writing using checklist and colla	Generating Organizing	–	–	+	Generating Organizing

Author, date	Experimental conditions	Planning	Feedback		Reviewing	
			Feedback by	Feedback on	Revising	Focus of revision on
	borative strategies					
	Cond 2: by means of video modeling in writing using checklist, and collaborative processes	Generating Organizing	–	–	+	Generating Organizing
	Cond 3: writing using checklist and collaborative processes	Generating Organizing	–	–	+	Generating Organizing
	Cond 4: collaborative processes	Organizing	–	–	–	–
7 Gunel, Hand, & Gunduz (2006)	Cond 1: Ppt	Organizing	Teacher	Organizing, Concept knowledge	–	–
	Cond 2: paper	Organizing	Teacher	Organizing Cconcept knowledge	–	–
	Cond 3: Ppt	Organizing	Younger students	Comprehensibility	–	–
	Cond 4: paper	Organizing	Younger students	Comprehensibility	–	–
9 Hand, Gunel, & Ulu, (2009)	Cond 1: math, math	Generating Organizing	Younger students, teacher	+	+	Comprehen- sibility, selecting, organizing
	Cond 2: math, graph	idem	idem	+	+	idem
	Cond 3: graph, graph	Idem	idem	+	+	idem
	Cond 4: graph, math	Idem	idem	+	+	idem

Author, date	Experimental conditions	Planning	Feedback		Reviewing	
			Feedback by	Feedback on	Revising	Focus of revision on
10 Hand, Hohenshell, & Prain (2004)	Cond 1: explanation,	Generating Organizing	Younger students	+	+	Comprehen-
	newspaper article	Generating Organizing	Newspaper editor	+	+	-sibility
	Cond: 2: explanation,	Generating	Self	+	+	Comprehen-
			Younger students	+	+	-sibility
	newspaper article	Generating	Peers	+	+	Organizing
			Newspaper editor	+	+	Comprehen-sibility
	Cond 3: explanation	Generating Organizing	Younger students	+	+	Organizing
	Cond 4: explanation	Generating	Self	+	+	Comprehen-sibility
			Younger students	+	+	Comprehen-sibility
						Organizing
						Comprehen-sibility
11 Hand, Hohenshell, & Prain (2007)	Cond 1a: textbook assignments	–	–	–	–	–
	Cond 1b: newspaper arrticle	–	Newspaper editor	+	+	Comprehen-sibility
	Cond 2a: explanation	Generating Organizing	Younger students	+	+	Comprehen-sibility
			Teacher	+	+	Concepts
	Cond 2b: newspaper article	–	Newspaper editor	+	+	Comprehen-sibility
	Cond 3a: newspaper article	–	Newspaper editor	+	+	Comprehen-sibility
	Cond 3b: explanation	Generating Organizing	Younger students	+	+	Comprehen-sibility
			Teacher	+	+	Concepts
	Cond 4a: textbook assignments	–	–	–	–	–
	Cond 4b: explanation	Generating Organizing	Younger students	+	+	Comprehen-sibility
			Teacher	+	+	Concepts
12 Hand, Wallace, & Yang (2004)	Cond 1:	Generating	–	–	-	
	Cond 2:	Generating	Peers	+	+	Comprehen-sibility
			Teacher	+	+	Comprehen-sibility
13 Hand, Yang, & Brux-voort (2007)	Cond 1: business letter	Generating	Younger students	+	+	Comprehen-sibility readability

Author, date	Experimental conditions	Planning	Feedback		Reviewing	
			Feedback by	Feedback on	Revising	Focus of revision on
14 Hohenshell & Hand (2006)	Cond 1:	Organizing Generating	Self	+	+	Representation terms, concep
	Cond 2:	Organizing Generating	Peers	+	+	Representation terms, concep
15 Hunter, & Tse (2013)	Text 1:	Generating Organizing	Teacher	Student's analysis of content in the future text	–	–
	Text 2:	Generating and organizing, both based on feedback on text 1			–	–
17 Kieft, Rijlaarsdam, & van den Berg (2006)	Cond 1:	Generating Organizing	Peers	–	–	–
	Cond 2:	Generating	Self	+	+	Organizing
			Peers	–	–	–
18 Kieft, Rijlaarsdam, & van den Berg (2008)	Cond 1:	Generating Organizing	Peers	–	–	–
	Cond 2:	Generating	Self	+	+	Organizing
			Peers	–	–	–
24 Mateos, Martín, Cuevas, Villalón, Martínez & González-Lamas (2018)	Cond 1: checklist (by teacher)	Generating Organizing	–	+	+	Generating, Organizing Content of argument
			\			
	Cond 2: self-study of check-list	Generating Organizing	–	+	+	Generating, Organizing Content of argument
35 Sampson, & Phelps Walker (2012)	Cond 1:	Generating Organizing	Peers	+	+	Goals, content of argument
Planning Only instruction						
16 Kabataş, Memiş & Öz (2017)	Cond 1:	Organizing	–	–	–	–
19 Klein, & Ehrhardt (2015)	Cond 1: stepwise goals	Generating Organizing	–	–	–	–
	Cond 2: clustered goals	Organizing	–	–	–	–
27 Nam, Choi, & Hand (2011)	Cond 1: sample A	Generating	–	–	–	–
	Cond 2: sample B	Generating	–	–	–	–

Author, date	Experimental conditions	Planning	Feedback		Reviewing	
			Feedback by	Feedback on	Revising	Focus of revision on
	Cond 3: sample C	Generating	–	–	–	–
28 Nevid, Pastva, & McClelland (2012)	Cond 1: relating chosen topic to learned subject matter	Organizing	–	–	–	–
	Cond 2: relating chosen topic to experiences	Organizing	–	–	–	–
29 Nevid, Pastva, & McClelland (2012)	Cond 1: relating assigned topic to chosen subject matter experiences.	Organizing	–	–	–	–
33 Rivard, & Straw (2000)	Cond 1: Sample: 11	Generating	–	–	–	–
34 Rivard (2004)	Cond 1: Sample: 39	Generating	–	–	–	–
41 Wissinger, & De La Paz (2016)	Cond 1:	Generating Organizing	–	–	–	–
43 Yildiz & Akdag (2021)	Cond 1: collaborative-ly story writing	Generating Organizing	–	–	–	–
	Cond 2: collaboratively column writing	Generating Organizing	–	–	–	–

[1]) *+ means: present;* ***-*** *means: not present; + in column* ***Feedback on*** *means: mentioned in* ***Focus of revision***

2.3.2 Effects of Forward Search instruction

As can be seen in Table 2.2, instruction in six studies is characterized as Forward Search. These studies took place in secondary education or university. When students receive Forward Search instruction, they are asked to write down all they can think of. Then, they receive feedback from a reviewer for revising their texts.

In three studies (1, 5, 8), reviewing comprises feedback and revising conforming to Klein's (1999) description of the Forward Search process, whereas in the remaining three studies (36, 40, 42), revising is postponed to a next writing assignment. In study 1, students are instructed to write answers on open-ended questions, then review each other's answers, and revise their own texts using the received feedback. Finally, students and teacher provide feedback in a classroom discussion, and if needed students revise their texts again. In studies 5 and 8, students are instructed to write to various types of audiences (to a non-expert audience, peers, parents, younger students and the teacher). These audiences provide feedback on comprehensibility of the drafts, which students use for revising.

In the other three studies (36, 40, 42), feedback is given in two ways: 1) a teacher provides written feedback on the application of, or reflection on course concepts; 2) a teacher as well as students provide feedback in classroom discussion, focusing on students' responses on prompts about general and specific aspects of a story (e.g.: What do you notice in the story? Or: How do you feel about Daisy?).

Of the six studies, there are three comparing Forward Search instruction to a business-as-usual control group (36) or a controlled control group (40 and 42). Two of these (36 and 42) report positive effects of Forward Search instruction. In study 36, two of four comparisons show positive effects on topic knowledge (large effect) and on a combined measure of insight and topic knowledge (medium effect). However, the other two comparisons regard a difference between honors and non-honors students, which cannot be considered as experimental comparisons of the intervention effects. Furthermore, initial differences between the (non-honors) experimental and the (non-honors) control group are not accounted for. Therefore, the difference in favor of the non-honors group receiving feedback on its writing may be attributable to other variables than the experimental intervention. In study 42, all four comparisons show positive effects of instruction on insight on two measures of insight (three large effect sizes, one medium). The third study (40) does not show any effect of the comparison on topic knowledge.

The remaining studies (1, 5 and 8) do not compare the results of the experimental conditions with a control group. The experimental conditions in study 1 and 5 show growth of insight, just as all four conditions in study 8.

Summarizing, three of six studies compare experimental conditions with a control group, and two of these (36 and 42) show positive effects on insight and topic knowledge, whereas the effects of study 36 may not be attributable to Forward Search instruction. Six of nine comparisons lead to positive results (four large and two medium effects). Additionally, study 1, 5 and 8 provide evidence that feedback by peers or various audiences followed by revision results in growth on insight.

2.3.3 Effects of Genre Writing instruction

Table 2.2 shows that in 14 studies, instruction is characterized as Genre Writing. These studies are directed at educational levels varying from grade 5 to university. In

all of these studies, instruction contains preparatory lessons on the characteristics of the genre in which students have to write.

The genres concerned are quite varied, such as essay, literature commentary, history synthesis, text with embedded modes, learning journal and narrative (see Table 1). According to Table 2, a number of these studies also contain instruction aiming at planning and/or revising in addition to genre knowledge, as follows. A group of three studies (3, 22 and 37) only focuses on the core business of Genre Writing instruction (genre knowledge). A second group consisting of six studies (2, 19, 30, 31, 38 and 39) contains planning instruction in addition to Genre Writing instruction. A third group consisting of five studies (21, 23, 25, 26 and 32) adds planning instruction combined with feedback and revision activities or adds feedback in combination with revision activities.

2.3.3.1 Plain Genre Writing instruction

Of the three studies focusing on genre knowledge, two studies (3, 22) compare one experimental condition with a control group. In study 3, students are instructed about literature by writing and discussing several genres (e.g., notes, minutes, synthesis). In comparison to a business-as-usual control group, the experimental condition achieves higher scores on insight, which are found on a posttest-writing task (literary commentary, large effect), but not on a measure of topic knowledge based on the same writing task. In study 22, instruction in the genre 'argument writing' is given by means of three consecutive steps: 1) modelling an argumentative text (teacher), 2) shared writing (teacher and students), and 3) guided writing (students, with the help of peers or teacher). This is an approach for language teaching called Genre Pedagogics (Rose, 2008). The experimental condition receiving this type of instruction does not outperform the control group on insight and topic knowledge in science (combined measure).

In study 37, two experimental conditions, on historical argument writing and non-disciplinary argument writing, are compared to each other on insight and topic knowledge. Both groups show growth in topic knowledge.

2.3.3.2 Genre Writing instruction complemented with planning

Four studies (20, 30, 31 and 38) compare experimental conditions with controlled control groups (20, 30 and 31) or a business-as-usual control group (38). In study 20, Genre Writing instruction takes place by means of Genre Pedagogics. This is complemented by instruction in planning containing inquiry activities (generating) and setting rhetorical goals (organizing). The experimental condition outperforms the control group on insight and topic knowledge in the field of science (one combined measure, large effect). In study 30, instruction in psychology classes is focused on explaining characteristics of learning journals, added with hints for writing (generating). This study reports positive effects on insight (one measure, medium effect) for the experimental condition, when compared to a control group after eight weeks. However, no effects are found on another, second measure of insight after 16 weeks.

Study 31 (a follow-up of 30) uses the same instruction and measures as study 30, but adds a second experimental condition in which the hints for planning are faded.

Fading starts halfway the first eight weeks. After eight weeks, the two experimental conditions, which are taken together, outperform the control group on insight (one measure, medium effect). However, after 16 weeks the faded experimental condition does not show a significant difference with the control group on the measure of insight. The condition receiving permanent hints even shows a significant decrease on insight after 16 weeks. In study 38, instruction consists of an explanation of learning journals combined with hints for generating (just as in studies 30 and 31). The experimental condition outperforms the control group on one direct measure of insight and two delayed measures of insight (three large effects).

Study 39 (a follow-up of 38) compares two experimental conditions: 1) hints for generating, and 2) the same instruction complemented with an explanation of the utility of writing assignments to students (as an extra motivator). The comparison shows that the motivator condition outperforms the other on a measure of insight (a writing assignment, large effect), but not on a combined measure.

Finally, study 2 (a case study) provides an instruction on essay writing, followed by students' discussion aimed at generating knowledge about the topic studied. In this study, growth in insight is measured by means of three consecutive writing tasks. These tasks prompt students to write about the same topic, from the angle of the same type of person (e.g., a farmer), but each time with a different aim: informing, expressing a feeling, and describing a dilemma. Growth in insight is found for 33% of the sample of 89 students.

2.3.3.3 Genre Writing instruction complemented with planning and/or reviewing

The third group consists of five studies (21, 23, 25, 26 and 32). These studies compare one experimental condition with a controlled control group. In study 21, Genre Pedagogics is used for instruction. This instruction complemented with planning and revising tasks results in negative effects on a combined measure for insight and topic knowledge in the field of science.

In study 23, model texts are used to instruct the experimental conditions in writing a synthesis of two source texts on history. Furthermore, students compose a checklist for self-assessing their texts on the characteristics of the genre synthesis after writing. Compared to a control group the experimental condition shows significantly higher scores on the combined measure for insight and topic knowledge (large effect). In study 25, the teacher explains a model text on chemistry containing embedded graphs and formulas and compares it to a text without graphs and formulas. After writing, students not only assess their texts themselves by means of an embeddedness checklist, but they also receive feedback from the audience they write for. In the post-test after the second unit, a positive effect is found for the experimental condition compared to a control group on a combined measure of insight and topic knowledge (medium effect). However, no effect is found in the posttest after the first unit.

Study 26 is a follow up of study 25. It applies the same type of instruction, but uses a larger sample. In comparison to a control group, it leads to positive results on one combined measure (medium effect). In study 32, the teacher models a narrative text about science. Furthermore, students discuss the contents of their texts (generating). After writing a first draft, students are instructed to give feedback to each other. This study reports a positive effect (small effect) for the experimental condition

compared to a control group on one combined measure for insight and topic knowledge.

To summarize the effects found for Genre Writing instruction, eleven studies (of 14) use experimental comparisons with control groups and nine of these show positive results. Eleven of seventeen comparisons show positive effects on insight and topic knowledge. The remaining three studies not comparing experimental conditions with a control group (2, 37 and 39) show evidence that writing texts about a same topic with different aims (2), using general as well as disciplinary argument writing instruction (37), and the use of a utility prompt as a motivator (39) may help increase insight and topic knowledge.

2.3.4 Effects of Backward Search instruction

As Table 2.2 demonstrates, instruction in 14 studies is characterized as Backward Search. All studies are directed at the upper half of secondary education and university. Backward Search comprises instruction in planning and reviewing activities and is applied in various ways. For planning, students are instructed to work with the Science Writing Heuristic in a number of studies. This heuristic requires students to perform laboratory activities, individually as well as collaboratively. The activities are aimed at selecting and organizing contents for students' texts. In some studies, students are required to plan modalities (formulas, graphs) in their texts. Concerning the reviewing activities, in nearly all studies students are instructed to revise their texts by using feedback from their audiences, which may be peers, younger students, teachers or the writers themselves. In two studies (7 and 15), writers receive feedback from their audiences, but do not revise their texts. However, in study 15, writers discuss the received feedback and use the outcomes for preparing a second text they are required to write just after finishing the reviewed draft.

Five studies (9, 12, 13, 14 and 15) contain a comparison between one or more experimental conditions receiving Backward Search instruction, and a controlled control group, or a business-as-usual control group. In studies 9, 12, 13 and 14, the control group works with an adapted program, and time on task is controlled. The four experimental conditions in Study 9 receive instruction to write two texts, and to embed one mode (math or graph) in each text, each condition using another sequence of modes. The study demonstrates that Backward Search instruction comprising instruction in the sequences 'math and next graphs', and 'math and next math' leads to positive effects on insight and topic knowledge when compared to a control group (medium and small effect). These sequences show larger effects than the other two investigated sequences.

In study 12, one experimental condition receiving Backward Search instruction, and a second experimental condition receiving instruction in planning only, are compared to a control group. In both experimental conditions, students collaborate using the Science Writing Heuristic for generating contents. Students in the Backward Search condition write to an audience, and revise their texts based on their audiences' feedback. The comparison to a control group is based on two measures (insight and topic knowledge). Results show positive effects on both measures for the Backward Search instruction condition (small effect on topic knowledge and large on insight), while instruction in Planning Only leads to positive effects on topic knowledge (small effect). In study 13, students are instructed to

discuss their ideas (generating) as a preparation for a writing task. Furthermore, students revise their texts with the help of their audiences' feedback. One of four comparisons with a control group shows a significant effect on insight (large effect), but not on topic knowledge. In study 14, students work with the Science Writing Heuristic in two conditions. In both conditions students write to an audience, a teacher or a peer, receive feedback and revise the text. This study has taken the two experimental conditions together for a comparison to a control group, which leads to positive effects on insight (large effect), but not on topic knowledge.

Study 15 compares one experimental condition to two business-as-usual control groups. Students in these groups complete the same writing tasks as the experimental condition, but do not receive Backward Search instruction. Results of the comparison to control group 1 after eight weeks show effects on insight (large effect) just as the comparison to control group 2 (small effect). Comparisons after four weeks do not yield positive effects.

Seven studies on Backward Search instruction do not contain comparisons with a control group, but compare different experimental conditions with each other (6, 7, 10, 11, 17, 18, 24) or are case studies without comparison (4, 35). In study 6, the number of instructional components declines from condition 1 to 4. Instruction in synthesis writing, collaborative learning and writing (the most components) shows the largest effects on insight. In four comparisons, study 7 shows that in Backward Search instruction, the mode of writing (power points vs. plain paper) makes a difference for learning: preparing power points resulted in positive effects on insight (two medium effects) and topic knowledge (large and medium effect). Studies 10 and 17 show that conditions containing organizing as part of planning outperform conditions in which text organization is the focus of feedback. In study 10, these positive effects are found on one combined measure and on two measures of insight (small, medium and large effects), while study 17 finds a positive effect on one measure of insight (small effect). Furthermore, study 10 shows that positive effects on insight (two measures) are larger after writing two texts, than after writing one text (medium and large effects).

Study 11 elaborates on the latter outcomes of study 10. Four classes write two texts in different genres. The students receive Backward Search instruction, instruction in reviewing or no-instruction in various sequences. After students have written one text, Backward Search instruction shows larger effects on insight than no-instruction on one measure (medium effect), whereas on another measure, no-instruction shows a larger effect on insight than Backward Search instruction (medium effect). When compared to no instruction, instruction in reviewing leads to positive effects on insight. When instruction in reviewing is used separately as well as combined with Backward Search instruction in two consecutive tasks, positive effects on insight are found on two separate measures (two medium effects) and on a total measure of insight (a combination of scores on three separate measures, medium effect), in comparison to no instruction.

In study 18, a replication of study 17, instruction in neither condition leads to positive effects on insight. Study 24 compares instruction in using a checklist for synthesis writing by means of modelling and a video with instruction for self-studying the checklist. Students in both conditions collaboratively write a synthesis. Instruction in using the checklist leads to larger effects on insight than self-study. Finally, in two case studies, students are instructed to work with a planning guide (4, 35), followed

by a peer review guide in study 35. The studies show that both types of Backward Search instruction lead to growth in insight (4, 35), and topic knowledge (35).

Summarizing, five studies (of 14) comparing Backward Search instruction with a control group provide evidence that there are positive effects on insight (four large, one small effect), on a combined measure (one medium, one small effect) and on topic knowledge (two small effects) for 50% of the experimental comparisons. Seven studies comparing different conditions for Backward Search instruction, additionally provide some evidence that elaborate instruction, the mode of writing (power points versus plain papers), the number of writing tasks and specific planning instructions (organizing vs. generating) are of importance for increasing insight and topic knowledge. In addition, two case studies provide evidence that Backward Search instruction may lead to growth in insight and topic knowledge.

2.3.5 Effects of Planning Only instruction

The fourth type of instruction in Table 2.2, Planning Only, is applied in nine studies. These studies are directed at grades 5 to 8, and at university level. Planning Only instruction is directed at planning activities in various ways. For instance, for generating, the Science Writing Heuristic and group discussions are applied, and for organizing, (non-textual) modalities, such as formula's or graphs.

Five studies compare one or more experimental groups with a control group. The control group receives business as usual (27, 41) or controlled instruction (16, 33 and 34). In study 27, three experimental groups are instructed to use the Science Writing Heuristic for generating a research question, first individually and later collaboratively. Two experimental groups show significant effects on a combined measure of insight and topic knowledge (both medium effects). In study 41, directed at generating content, experimental students are instructed to discuss collecting arguments from historical sources. The teacher uses an argumentation scheme and critical questions to stimulate students' discussion. The experimental group shows a large effect on a measure of topic knowledge.

In study 16, experimental students are instructed in embedding formula's and graphs for organizing text contents, whereas control students carry out the writing task, without using such non-textual modalities. This Planning Only instruction leads to a large effect on one measure of insight. Study 33 instructs experimental students in generating by discussion in small groups. Comparisons to a control group performing other tasks (e.g., 'fill in the blanks)', show that generating by discussion leads to effects on two measures (direct and delayed) of insight (medium and large effects respectively) and two measures (direct and delayed) of topic knowledge (large effects). However, because of the small sample (n = 11), the positive effects have to be treated with caution. Study 34, a follow-up of study 33, uses the same intervention in a larger sample (n = 39). In this case, experimental comparisons do not result in significant effects on two measures of topic knowledge (direct and delayed).

In two studies (19, 43), two experimental conditions are compared to each other. Study 19 compares a clustered instruction providing directions for organizing subgoals with instruction guiding students stepwise through generating and organizing goals for writing. No significant differences between the two conditions are found on a measure of topic knowledge. In study 43, students in two experimental conditions are instructed how to plan and write a text for peers collaboratively, in one

condition they have to write a story, in the other a column. The study shows a medium effect on students' insight in favor of writing a story compared to writing a column.

The studies 28 and 29, of which 29 is a follow up of 28, do not contain comparisons with another condition. Instruction is directed at organizing. Study 28 requires students to choose topics from a list. In study 29, students are instructed to choose between two genres. Both studies show positive effects on students' knowledge of topics they have written about, on one measure.

Summarizing, in five studies (of nine), Planning Only instruction shows positive effects on insight (two large effects, one medium effect), on topic knowledge (three large effects) and on a combined measure of insight and topic knowledge (two medium effects), in comparison to a control group. In total, 73% of the experimental comparisons to a control group shows positive effects. One study comparing two experimental conditions on insight shows that the genre of the writing task (story vs column) may determine learning (medium effect). Two studies without comparison groups provide some evidence that making a choice between topics (28) or between two genres of writing (29) may enhance topic knowledge.

2.4 Conclusions

Table 2.3 compares the experimental effects of the four types of instruction. In the second column, for each type of instruction the number of studies comparing experimental to control groups is shown. In the next column, the total number of comparisons of experimental to control groups in these studies is presented, followed by a column showing the number of comparisons with positive effects. The percentages in the fifth column are computed by dividing the number of comparisons with positive effects by the total number of comparisons. The remaining three columns give the numbers of positive effects for each of the three types of posttests that are used in the studies and their effect sizes. The comparisons are used in 24 studies. Table 3 involves experimental comparisons only, because these can provide a fairly strong basis for formulating conclusions about effects of the four types of instruction.

It appears from Table 2.3 that all four types of instruction may lead to positive, mainly large and medium effects on insight and topic knowledge. However, in about one third of the cases, there are no positive effects on learning. The number of studies showing effects varies per type of instruction. The results of Forward Search are based on a relatively small number of studies (3). Therefore, this type of instruction can be considered as weakly supported by experimental evidence.

When the results of the three types of Genre Writing instruction are taken together, the percentage of positive effects on learning is 65% (six large, four medium and one small effect size), which is nearly two third of the experimental comparisons. Given the number of studies (11), the support for Genre Writing instruction as a means to stimulate learning is substantial. In eleven studies on Genre Writing instruction, eleven of seventeen comparisons with a control group lead to positive effects on insight and combined measures of insight and topic knowledge. Of the three types of Genre Writing, Genre Writing instruction complemented with planning, and Genre

Writing complemented with planning and/or reviewing lead to more positive effects on insight and topic knowledge than instruction in genre knowledge only.

According to Table 2.3, five experimental studies testing Backward Search instruction, show that nine out of 18 comparisons with a control group (50%) lead to positive effects on insight and/or topic knowledge (four large, one medium, four small effects). This means that experimental evidence for Backward Search instruction for stimulating writing-to-learn, is rather weak.

Table 2.3 shows that five studies applying Planning Only instruction comprise eleven comparisons to a control group, of which eight lead to positive effects on insight and/or topic knowledge (five large, three medium effects). Although the percentage of positive effects on learning (73%) is quite high, the number of experimental studies comparing to a control group is relatively small (5). Therefore, we regard support for Planning Only instruction for writing-to-learn as inconclusive.

Additional to the results of the experimental comparisons, results of case studies and studies comparing experimental conditions are of importance, because these studies provide alternative options for how to stimulate reflection by instruction, as described in the Results section. Examples are: using reviewing twice in Forward Search, or adding specific elements to instruction, such as a motivator (utility prompt) in Genre Writing.

Regarding the disciplines involved in the reviewed studies, it appears that writing-to-learn activities can be applied in a large variety of disciplines. The largest part of the studies (23) is directed at science, comprising seven different disciplines. The remaining studies (20) are performed in seven other disciplines belonging to humanities, social studies, behavioral sciences and earth sciences. Concerning the grades, most studies are directed at secondary education (grade 7-12). Much less studies are conducted in higher education, most of which with undergraduate students. The number of studies found in primary education is the smallest (grades 5 and 6).

In the present review, we found that we can classify instruction in studies on writing-to-learn by means of the four types of instruction that have been distinguished. Future research can use these types for characterizing and comparing treatments directed at writing-to-learn. This may improve the theoretical and practical use of the distinction between the four types of instruction. In the discussion section below, this idea is elaborated.

Table 2.3 Positive effects of four types of instruction in experimental comparisons with control groups

Type of instruction	Number of studies	Number of comparisons with control group	Number of comparisons with positive effects	Percentage of comparisons with positive effects	On combined measure of insight and topic knowledge	On separate measure of insight	On separate measure of topic knowledge
Forward Search	3 (of 6*)	9	6	67 %	1L**	3L 1M	1M
Genre Writing	11 (of 14*)	17	11	65 %	2L 2M 1S	4L 2M	–
Backward Search	5 (of 14*)	18	9	50 %	1M 1S	4L 1S	2S
Planning Only	5 (of 9*)	11	8	73 %	2M	2L 1M	3L

* *all reviewed studies applying this type of instruction* * * *Small (S), Medium (M) and Large (L) effect sizes*

2.5 Discussion

The results of the present study provide evidence for positive effects of the four types of instruction on learning by writing, though not to the same extent, as explained above. The results must be considered with caution, because the number of studies for each type is quite small and not all designs are equally strong. As can be seen in Table 2.1, the sample sizes of the 24 (quasi) experimental studies vary between 32 and 1031 participants, giving quite different weights to each of the studies. Another methodological difference is whether randomization is used (once in the three studies on Forward Search, eight times in the eleven studies on Genre Writing, twice in five studies on Backward Search, and four times in five studies on Planning Only). In three of the latter, individual randomization was applied, just as in two studies on Genre Writing. The remaining random assignments were on the level of the group. Additionally, all of the reviewed studies took place in regular classrooms and therefore are dependent on practical issues, for instance: changes in classroom scheduling, rules for testing, or dropout of students and teachers. Such events may have had an influence on the integrity of the designs and results of the reviewed studies.

2.5.1 Mechanisms underlying writing-to-learn

In this section, we provide explanations for our conclusions about the effects of the four types of instruction, by hypothesizing about the mechanisms underlying the writing-to-learn process.

2.5.2 Mechanisms underlying Forward Search instruction

Forward Search instruction requires students to write down all ideas they can think of, and after finishing a first draft, feedback is provided and instruction on revising is given. Galbraith (1999) considers the writing of a first draft as externalizing students' knowledge as it is represented in their mind, providing the possibility to reread and reconsider one's own knowledge. This written display of knowledge is supposed to initiate a loop of rereading, feedback and revising directed at recognizing and acquiring new insights into conceptual relations and at accommodating to rhetorical demands.

To stimulate this constitution of new knowledge as Galbraith (1999) calls it, feedback followed by revision is brought into the writing process by instruction. Klein (1999) and Galbraith and Torrance (2004) consider revision crucial for acquiring knowledge. Expert writers' focus for revision is on the meaning of their texts and on coherence (Klein, 1999), or on identifying potential new insights (Galbraith, & Torrance, 2004). Foci for feedback in Forward Search instruction can be derived from these descriptions. Such foci are relating concepts, making inferences, organizing content and identifying new insights. These foci stimulate writers to reread their draft critically by asking themselves questions (for instance: did I draw the right inference?) and to revise it.

Wallace et al. (2007) propose that feedback should also pay attention to rhetorical aspects of texts. They argue that writing should be directed at an audience, which has consequences for how ideas are being formulated. When writing for their teacher, students may just write what they know, but writing for an audience requires them to view their topic from their audience's perspective. Then, they have to ask questions such as: what knowledge does my audience have? Which information is appropriate? When revising using feedback focused on such rhetorical aspects, students have to think about how to formulate their ideas. This reformulation may lead to a new perspective and may therefore result in new insights into subject matter (Prain, 2006).

The assumed mechanisms of Forward Search may have been activated in three studies, because three of six studies which contain a comparison to a control group (six out of nine comparisons) provide positive evidence for the idea that Forward Search instruction leads to learning. Thus, the evidence on the effects of Forward Search instruction on writing-to-learn with only six empirical studies is quite meagre. Therefore, more experimental studies are required to decide whether this type of instruction is generally effective in stimulating the writing-to-learn process.

2.5.3 Mechanisms underlying Genre Writing instruction

Genre Writing instruction consists of an explanation of the nature of a genre as preparation for writing. According to Newell (1984) writers select the rhetorical goal belonging to the genre in which they are writing. Furthermore, they use the genre structure for constructing relations between ideas and for selecting content. The constraints that a genre poses force writers to rethink the order of their ideas and the relations between them, and to look at them in new ways, which may result in new insights into the topic of their writing.

For writing in a particular genre, writers need specific genre knowledge. This means that students have to understand the macro-structure of the genre, but they also have to know what type of vocabulary and sentence structure can be used. Because it is not self-evident that students dispose of sufficient genre knowledge (Schleppegrell, 2004), studies classified as Genre Writing instruction provide genre knowledge in preparatory lessons to scaffold students before entering the writing process. Additional support can be given by instruction on planning using the structure of the genre as a coat rack, and by instruction on revising based on feedback that is focused on specific genre characteristics, such as vocabulary, register and sentence structure.

The Genre Writing studies encountered in our review were of three types: 'plain' Genre Writing instruction directed at genre knowledge in advance of writing, Genre Writing added with instruction in planning, and Genre Writing added with instruction in planning and/or reviewing. Results show that the second and third type led to more positive effects on insight and topic knowledge than 'plain' Genre Writing instruction. It therefore seems that additional instruction directed at planning or additional instruction directed at planning and/or revising reinforces the effects of 'plain' Genre Writing instruction on students' insight and topic knowledge. The number of experimental studies directed at Genre Writing instruction using a control group (11) is large in comparison to the other types of instruction in our review. Therefore, the positive evidence that Genre Writing instruction (especially the second

and third type) is effective in inciting the process of writing-to-learn in students is relatively strong.

2.5.4 Mechanisms underlying Backward Search instruction

Backward search instruction consists of instruction in planning as well as reviewing. When reviewing, writers compare their contents to their rhetorical goals and ask themselves whether they have succeeded in their original (planned) intentions.

Klein (1999) bases his Backward Search hypothesis on the model for expert writing of Bereiter and Scardamalia (1987). According to this model, writing is a knowledge transforming process, in which new insights arise. Students set rhetorical goals, which serve the interest and knowledge of their audiences. In addition, they generate content from their conceptual knowledge or external sources. There is a constant exchange of ideas between the content space and rhetorical space in order to arrive at a better fit between the two. When writers realize that their ideas from the content space and the rhetorical space do not match, they will adjust their ideas from the content space, or they will reconsider their rhetorical goals. This matching process leads to new insights and topic knowledge. In accordance with this theoretical mechanism, Backward Search instruction demands from students to perform planning as well as revising activities based on feedback. This sequence of planning, formulating, feedback and revising in Backward Search instruction is intended to stimulate the recursive interaction between rhetorical and content space while writing. In our review, nine of 18 experimental comparisons to control groups in five studies show positive effects on insight and topic knowledge, or on insight only. We, therefore, have to conclude that empirical support for the beneficial effects of Backward Search instruction on learning is not convincing. There are relatively few studies allowing comparisons with a control group, and only half of these comparisons show positive effects on topic knowledge or insight.

The remainder of the studies into effects of Backward Search instruction consists of comparisons between experimental conditions (seven studies) and of small case studies (two). Although these studies are useful for optimizing conditions for Backward Search instruction, such as the type of audience providing feedback or the focus of feedback provided (see Table 2.2), they are not suited for evaluating the effects of Backward Search instruction and therefore do not provide evidence for the validity of the hypothesized process underlying writing-to-learn. Given the few studies providing evidence on the effects of Backward Search instruction, there is a need for future studies containing experimental comparisons of Backward Search instruction to control groups on their effect on students' insight and topic knowledge.

2.5.5 Mechanisms underlying Planning Only instruction

Planning Only instruction entails instruction on planning activities only. Therefore, it can be seen as the opposite of Forward Search instruction that focuses on revision only. Langer and Applebee (1987) state that manipulating contents in various ways will contribute to learning. Wallace et al. (2007) regard the performance of various planning activities by the writer as a way to arrive at insight. These activities may, for instance, be weighing which contents will be part of the text, exchanging ideas with other persons and making notes to organize contents. Planning activities may also include students comparing their own selection of contents with their peers', which

may urge them to adapt their own. By carrying out these types of activities, students may reconsider their selection of content elements for their text to be written. When adapting and reconsidering their planning, new insights and topic knowledge may arise. Galbraith (2015) also states that planning activities may lead to new insights, in case writers negotiate contents intensively.

In the reviewed studies, Planning Only instruction is applied by providing students with a checklist (Science Writing Heuristic) for performing various laboratory activities, or by demanding students to plan in peer groups by brainstorming, elaborating and evaluating their ideas. Eight of eleven experimental comparisons to control groups for Planning Only instruction show positive effects on insight and topic knowledge. This may seem quite substantial support for the idea that Planning Only instruction leads to adapting and reconsidering contents and therefore results in writing-to-learn. However, given the fact that only five experimental studies are involved, more experimental research is needed for evaluating the effects of this type of writing instruction.

2.6 Suggestions for future research

We have to consider the fact that research into instruction directed at writing-to-learn is of a quite recent date. As remarked previously, it was only after Klein's (1999) seminal article about cognitive processes underlying learning by writing that research turned to the types of instruction that are needed for triggering this process with students of different ages and courses. Therefore, it is not surprising that empirical evidence about the effectiveness of the different types of instruction is still inconclusive. At present, conclusions about the effects of the four types of instruction on learning should be drawn with great care, as results so far are quite mixed and based on relatively few studies that offer hard experimental evidence.

Even for Genre Writing instruction, which is the most studied type of the four distinguished, not more than eleven studies offer experimental comparisons with control groups, taking into account that this group of studies is diversified over subgroups (with planning, or with planning and/or revision). For that reason, it is important that future studies are carried out into each of the types of instruction using strong experimental designs and with a strong theoretical basis, elucidating the assumed underlying processes of writing-to-learn. For example, what is the role of feedback in these processes and how is that role enacted in the classroom context?

The role of instruction in writing-to-learn in the disciplines is central in this review. Therefore, we selected studies embedded in classroom contexts. This selection resulted in a large number of different disciplines, for instance: science, humanities, social and behavioral studies. However, some disciplines are missing. Most striking is the absence of mathematics. Although writing-to-learn studies are conducted in mathematics, they do not meet our criteria for inclusion. Either they are conducted before 1999 (Crocker, 1992; Kasparek, 1996), or they are not focused on effects of writing-to-learn instruction on students' insight and topic knowledge, but for instance on the teachers' skills to implement writing-to-learn in math class (Akkus, & Hand, 2011; Kenney et al., 2014; Diaz Eaton & Wade, 2014). Why are effects of instruction in writing-to-learn barely studied in mathematics? Our explanation is that

this may be related to the attitude of teachers in this discipline. Linnemann and Stephany (2014) mention that math teachers hardly reflect on the possible yield of applying writing for learning in their teaching practice, and therefore do not use writing assignments. Furthermore, Brozo and Crain (2018) state that mathematics teachers do not consider writing as relevant for mathematics learning, while Teuscher et al. (2015) report that many math teachers view writing-to-learn activities as consuming too much time. These observations may also be an explanation for researchers' focus on the teacher in studies on mathematics. We suggest research into the question whether instruction in writing-to-learn in mathematics classes leads to positive effects on learning. It is worthwhile to find out whether writing-to-learn in this discipline is an effective way of learning.

Evidence for the process of writing-to-learn is shown only indirectly in the reviewed studies. When instruction leads to more insight and topic knowledge, a mediating effect of the process of writing-to-learn is assumed. Above, we discussed the mechanisms underlying the process of writing-to-learn and determined whether the presence of these mechanisms can be deduced from the found effects of the four types of instruction. However, more insight into the occurrence of these mechanisms is needed. Think aloud studies on the thinking process of individual students while carrying out writing-to-learn tasks under different conditions may offer interesting clues about how learning of topic knowledge and acquiring new insight is brought about by the cognitive processes involved in writing. More of this type of scientific knowledge about the process of writing-to-learn is a valuable source for improving instruction in the use of writing as a learning tool.

2.7 Implications for education

The present review discriminates four types of instruction (Forward Search, Genre Writing, Backward Search and Planning Only), all of which may result in increased insight and topic knowledge of students. Although evidence is not as strong as we may wish for, we will explore the question how knowledge acquired in our review can be implemented in education. The reviewed studies show various ways to apply the types of instruction, such as working with a checklist for planning, or giving feedback on comprehensibility for an audience of readers. These types of instruction can be applied in education as an alternative for more traditional approaches to learning. When teachers understand the basic assumptions underlying each of the four instructional approaches (the so-called mechanisms of writing-to-learn), they may experiment in their classes with writing assignments and instructional support. Then, teachers can find out how the various types function in class. Some students, for instance low self-monitors, may be better served with free writing, guided by no more than a subject to write about (Forward Search instruction), for activating their prior knowledge. Other students may need structured assignments and benefit more from precise direction of their writing process (which is the case with Genre Writing, Backward Search and Planning Only instruction), for instance when writing a lab report.

Not every teacher interested in writing-to-learn, and understanding the mechanisms, will find it easy to provide good instructions. Baker et al. (2008) report

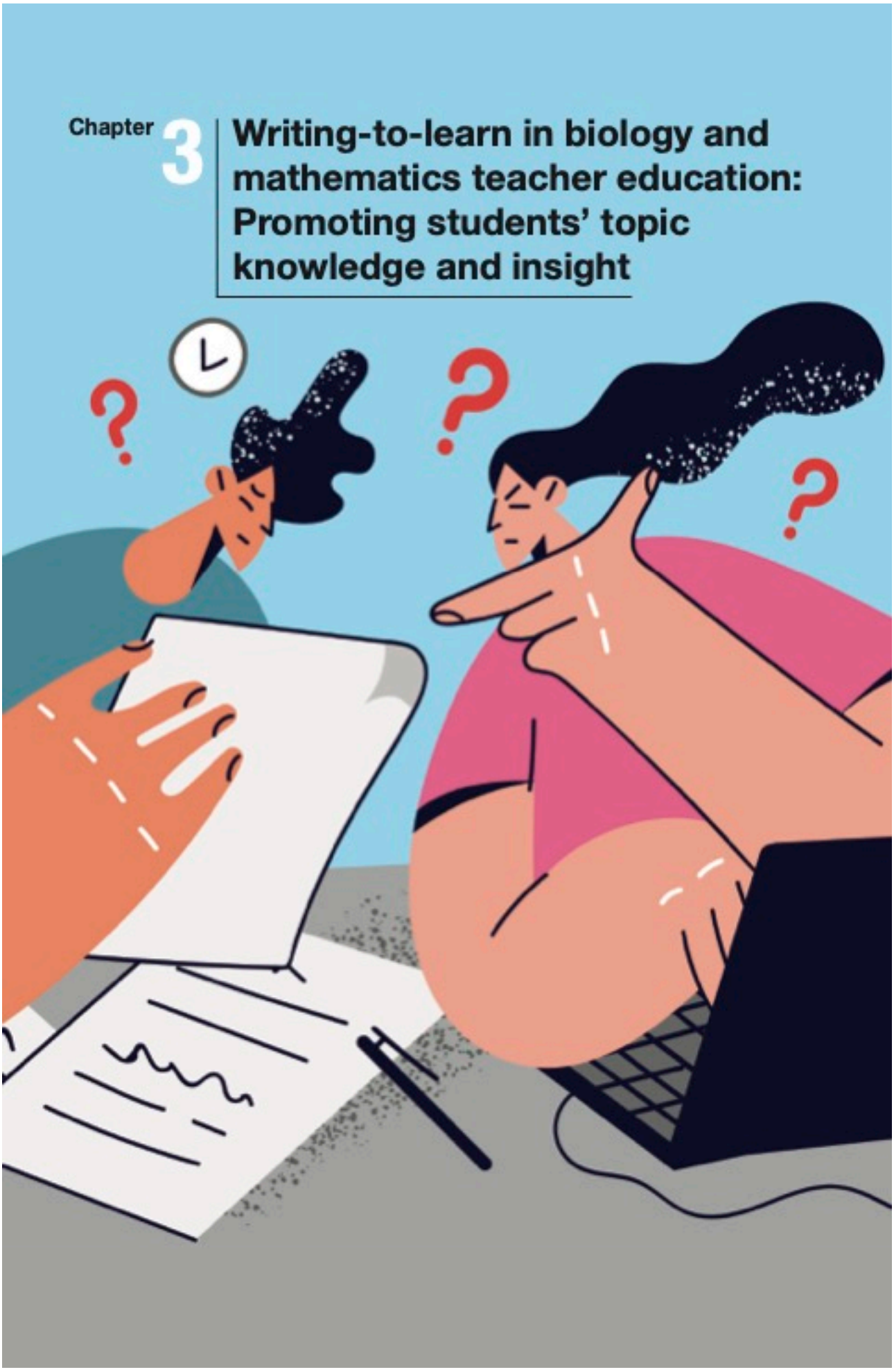
teachers' problems with the implementation of writing-to-learn in class. In interviews with the researchers, these (science) teachers state that they are no language teachers and do not know how to evaluate students' texts and how to give them feedback. Furthermore, they are concerned about the amount of time writing-to-learn needs.
When asked for solutions, interviewees suggest that these teachers in disciplines such as science or social studies work together with language art teachers for experimenting with writing-to-learn in their programs. Such teams can discuss the design of writing tasks, and appropriate types of instruction (Baker et al., 2008). Furthermore, they can compose writing assignments including instruction. Language art teachers can explain that students need concrete directions for writing, meaning that they need to know the goals and the audience they write for, as well as the genre of the text to be written. Instruction in planning (Backward Search, Genre Writing) can be facilitated by choosing an existing checklist (such as the Science Writing Heuristic), of which this review shows positive effects on learning. Implementing Genre Writing instruction requires collecting or designing appropriate model texts. Language art teachers can propose criteria for such texts (e.g., text difficulty, vocabulary, text structure) and provide concrete examples of model texts in a genre (e.g., explanations, arguments, journals etc.).

Teachers will have to prepare for coaching the writing process, particularly for giving feedback. Bean (2011) suggests that the application of a web-based reciprocal peer review system stimulates students' self-support and saves time. Students upload their texts in the system, give feedback to each other, and evaluate their peers' feedback, by means of rubrics. In the end, teachers can add their feedback. The language art teachers and teachers of other subjects can compose the rubrics. These should be compatible with the goals of the writing task, directed at topic knowledge of and insight into subject matter. This type of experimenting with writing-to-learn by collaborating in teams may give teachers more confidence in applying this new way of learning in class.

Positive effects of the four types of instruction that are found in this review give a glimpse of the hypothetical underlying mechanisms of writing-to-learn we have described. This is a promising start, but more experimental evidence for how different types of instruction stimulate the process of writing-to-learn is needed, because the number of studies providing evidence hitherto is small.

Chapter 3

Writing-to-learn in biology and mathematics teacher education: Promoting students' topic knowledge and insight



Chapter 3

Writing-to-learn in biology and mathematics teacher education: Promoting students' topic knowledge and insight[2]

Abstract

In the present study, effects of Genre Writing instruction added with planning and revising activities (GWPR) are investigated in teacher education. This type of instruction was considered promising because it appeared to lead to positive effects on topic knowledge and insight in previous studies conducted in secondary education. Researchers' expectation was that writing-to-learn activities by means of GWPR support teacher candidates in acquiring topic knowledge and insight into subject matter. Two studies were undertaken, one in biology and one in mathematics teacher education, each comprising a quasi-experiment and a think aloud study. Both studies were embedded in regular courses. Researchers co-created writing-to-learn tasks with the teacher educators involved. Both experiments showed positive effects on learning. Results of the think aloud studies provided evidence for specific indicators (students' reflections) of the process of writing-to-learn, in which experimental teacher candidates differed from the control group.

[2] This chapter is a slightly modified version of the article:
Van Dijk, A., Van Gelderen, A., & Kuiken, F., (2023). Writing-to-learn in biology and mathematics teacher education. Promoting students' topic knowledge and insight. *Frontiers in Education,8.* |https://doi.org/10.3389/feduc.2023.1094156

3.1 Introduction

Students in higher education often have difficulty understanding the contents of their textbooks and their teacher educators' explanations of subject matter (Hunter & Tse, 2013). Consequently, they cannot acquire the topic knowledge and insight a course is aimed at. Students' unfamiliarity with discipline specific language in academic genres is seen as an important cause of their difficulties with learning (Hunter & Tse, 2013). Several studies emphasize the importance of supporting students in acquiring the desired topic knowledge and insight (Sampson & Phelps Walker, 2012; Hunter & Tse, 2013; Finkenstaedt-Quinn et al., 2017). By topic knowledge we mean basic factual knowledge in the context of academic courses. By insight we mean the ability to relate new concepts to students' prior knowledge. Insight is viewed as the ultimate aim of academic courses, because it is the manifestation of a higher order understanding of concepts (Boscolo & Carotti, 2003; Hand et al, 2009, Klein & Rose, 2010, Kenney et al., 2014).

Teachers have different means at their disposal for providing support, for instance by teaching task-oriented reading, taking into account students' limited genre specific vocabulary and grammar knowledge, or by stimulating students' reflection about subject matter. Another way of supporting students is using writing as a tool for learning. This means, that students carry out writing tasks that are intended to stimulate reflection on their writing resulting in new topic knowledge and insight, better known as writing-to-learn (Galbraith, 2009; Bereiter & Scardamalia, 1987). Examples of writing-to-learn tasks are writing a learning journal for reflection on subject matter or writing about discipline specific concepts adapted to a lay audience. Writing is considered an important medium for learning, because externalizing reflection in a written text enables writers to retain their thoughts and allows them to reread and develop their ideas further (Nückles et al., 2009). This may lead to new topic knowledge and insight into disciplinary concepts (Emig, 1977; Klein & Boscolo, 2016; Graham, et al., 2020).

The present study aims in the first place at the co-construction of instructional materials and assessment of the effects of a method for writing-to-learn directed at students in higher education. The method is based upon previous research into instruction on writing-to-learn. A second aim is finding characteristics of the process of writing-to-learn for a better understanding of how it operates.

3.2 Research into instruction in writing-to-learn

Although writing-to-learn has been studied for several decades and has shown positive effects on learning, it is not much used by teachers (Klein & Boscolo, 2016). One often- mentioned reason is that many teachers do not know how to implement writing-to-learn in their curriculum (Akkus & Hand, 2011; Kenney et al., 2014; Diaz Eaton &

Wade, 2014). Research into promising approaches for writing-to-learn has given ground for recommendations for teaching practice. Below, we give a brief overview. In the first meta-analysis on writing-to-learn carried out (48 studies), Bangert-Drowns et al. (2004) concluded that in most studies involved, writing was not accompanied by specific instruction directed at writing-to-learn and did not result in significant learning outcomes. However, they found a minority of studies in which instruction in metacognitive writing strategies was provided, which led to positive results. Bangert-Drowns et al. (2004) speculated that instruction in cognitive and metacognitive writing strategies is promising for future research into effects of writing-to-learn, because these types of instruction stimulate reflection on writing products, which may lead to learning of course contents. Cognitive writing strategies are understood as organizing strategies, such as goal setting, selecting and structuring contents (Berthold et al., 2007). Metacognitive writing strategies are understood as strategies for monitoring own task performance and evaluating texts (reviewing and revising) (Berthold et al., 2007).

In their meta-analysis of 56 experiments including 19 studies examined by Bangert-Drowns et al. (2004), Graham et al. (2020) analyzed effects of writing-to-learn activities on learning in science, social studies and mathematics in grades 1-12. They found a positive effect with an average effect size of .30, (which was larger than the findings of Bangert-Drowns et al. (2004), who found an average effect size of .17). However, there was a wide variability in effect sizes found, ranging from 1.67 to –0.74. Eighteen percent of the experiments showed negative effects. The authors could not explain this variability with any of a large group of variables, such as the type and features of activities and instruction (including the use of cognitive and metacognitive strategies), study characteristics, discipline (science, social studies or math), grade, duration et cetera. Moderation analyses using each of these variables showed no significant results. The authors concluded that the descriptions of most studies were not sufficiently detailed to determine which of the contextual and instructional factors were actually involved, for instance the details of writing tasks used, or the type of thinking operations instruction was meant to provoke (directed at topic knowledge or insight). Therefore, Graham et al. (2020) call for much more detailed descriptions of the contextual and instructional features of writing-to-learn interventions in future studies.

Miller et al. (2018) conducted an inductive, systematic literature review of 43 studies, embedded in regular courses in grades 6-12 in science, social studies and mathematics. They investigated the state of research on the use of writing-to-learn tasks in content areas by studying effects of instruction on learning. The researchers distinguished explicit instruction, inquiry-based instruction of cognitive and metacognitive strategies, and instruction of self-reflection. Explicit instruction can be provided by means of a model or a checklist, or by directly instructing the planning of a writing task. Inquiry-based instruction stimulates students to find out how to use cognitive and metacognitive strategies without guidance by the teacher. Instruction in self-reflection entails journaling requiring students' reflection on their own learning. Miller et al. (2018) concluded that both explicit and inquiry-based instruction of cognitive and metacognitive strategies as well as instruction of self-reflection can be effective in stimulating learning by writing. Overall, in 46,5 % of the reviewed studies,

instruction of metacognitive and cognitive strategies, and self-reflection clearly promoted learning.

Van Dijk et al. (2022a) reviewed 43 studies (from grade 5 to higher education) investigating which types of instruction of cognitive and metacognitive strategies lead to topic knowledge and insight. In this study, four types of instruction for writing-to-learn were discriminated. The first three were based on Klein's (1999) hypotheses about the nature of cognitive operations involved in writing-to-learn. The fourth type emerged from a number of reviewed studies. The first type that was called Forward Search stimulates the use of metacognitive strategies for reflection on contents of a draft. The second type that was called Backward Search requires the use of cognitive as well as metacognitive strategies. Cognitive strategies are directed at setting goals and planning text contents, and metacognitive strategies are directed at revising a draft referring to a previously made plan. The third type that was called Genre Writing consists of the provision of genre knowledge added with cognitive and/or metacognitive strategies for planning and revision. The fourth type that was called Planning Only consists of cognitive strategies for planning. The review found positive effects on topic knowledge and insight for all four types in 62% of experimental comparisons on average. The most empirical support, given the number of studies, was found for Genre Writing if it was supported with additional instruction in planning and revision.

Other studies point at elements of instruction that may enlarge effects of strategy-instruction: the intended audience and genre knowledge. Prain (2006) suggested requiring students to write for a lay audience (an audience unfamiliar with the topic). Presumably, this urges students to reflect on formulations matching their audience's knowledge, leading to a critical review of their insights. Prain's (2006) suggestion was tested in various studies (Finkenstaed et al., 2017; Hand et al., 2007; Hohenshell & Hand, 2006; Hand et al., 2004) showing positive results. Newell (1984) emphasized the role of genre knowledge for writing-to-learn. If students have knowledge of the requirements of the specific genre they are supposed to follow, they may be able to recognize relations between concepts they were not aware of previously (Newell, 1984). Klein & Kirkpatrick (2010) and Klein & Samuels (2010) confirmed Newell's (1984) view on the role of genre knowledge by showing that the effect of genre writing instruction on learning may depend on students' knowledge of the genre. Klein & Boscolo (2016) noted that genre knowledge may be a prerequisite for writing-to-learn.

Although most empirical studies mentioned above provide evidence for positive effects of instruction in genre knowledge, metacognitive and cognitive strategies, and writing for a lay audience on learning, evidence is still inconclusive, given that for each type of instruction null results of experimental comparisons are quite frequently encountered (Miller et al., 2018: 53,5 %, Graham et al., 2020: 19%; Van Dijk et al., 2022a: 38 %).

3.3 Promising elements of instruction in writing-to-learn

In this study, the evidence for the elements genre knowledge, planning and revising (cognitive and metacognitive strategies) directed at a lay audience as providing positive effects, is followed as a lead. The combination of these elements is what we call 'Genre Writing instruction added with Planning and Revising' (GWPR). Studies using this combination of instructional elements in secondary education were among the most successful studies (Van Dijk et al., 2022a). This study sets out to find more conclusive evidence for the effects of GWPR on disciplinary topic knowledge and insight of students in higher education.

In the first place, GWPR instructs genre knowledge in a preparatory activity (before the actual execution of a writing assignment) by means of a model text of the genre at stake. Genre knowledge is defined as knowledge of the genre's rhetorical goal and prevalent conceptual relations between text elements to arrive at this goal (Halliday & Martin, 1993). For instance, the rhetorical goal of explanatory texts is clarifying a topic, and a prevalent conceptual relation, for instance 'comparing', shows the disciplinary use of the genre (e.g., the comparison between ecological niches of different species in biology, or the comparison of numerical equations in mathematics). Analyzing a model text can make students aware of these characteristics. The model text should therefore exemplify various linguistic realizations of the conceptual relation, for instance the relation 'comparing' is realized as '... differs from…', 'more recently …'or '… is larger than…' (dependent on the specific disciplinary context). If students are made aware of how the conceptual relation 'comparing' in an explanatory text can be realized, they can make comparisons in their own writing and reflect on the results, which may lead to new insights in the meaning of these comparisons (Langer & Applebee, 1987). In addition, in GWPR, the model text is written in such a way that it is comprehensible to a lay audience. Therefore, no disciplinary jargon is used.

In the second place, planning consists of the cognitive strategies selecting and organizing contents in preparation of writing. In our view of GWPR, instruction on planning entails that students can use pairwise brainstorming aimed at the selection of knowledge elements from memory and textbook that they find relevant. For instance, in case of a text about similarities and differences between old and contemporary views, students will have to decide which elements of these views are relevant for such a comparison. Students are instructed to represent their selection by means of keywords and to organize them in a mind map, such that the structure of their draft becomes visible. They thereby have to consider the conceptual relations in view of comprehensibility for their audience and may therefore decide to include an introduction or a conclusion. While writing their drafts (individually) they are supposed to consult their planning as well as the model text exemplifying the formulation of central conceptual relations.

In the third place, in GWPR, revising consists of the use of metacognitive strategies for reformulation on the basis of peer feedback. Students are instructed to review their peer's draft focusing on the conceptual structure as realized in the text and on its comprehensibility for a lay audience. For instance, if the conceptual structure is based on making comparisons, peers check the clarity and accuracy of the

comparisons made, and whether they are in accordance with the writing assignment. In this process, students reflect on their peer's representation of the conceptual relations in language from the viewpoint of the intended (lay) audience. Students use their peer's feedback for revising their drafts individually. In doing so, they have to reflect on their original insights into the conceptual relations and their original formulation, which may lead to new insights (Bereiter & Scardamalia, 1987).
In the fourth place, writing for a lay audience is used as intrinsic in the planning and revising phases of GWPR instruction. Additionally, a lay audience is an important condition for finding appropriate sources for model texts. It is defined as an audience that is not familiar with the disciplinary contents of the course. It may consist of younger students than the writers or of a general audience. When planning and revising a text for a lay audience, writers cannot copy disciplinary jargon from their textbooks. Students therefore have to reflect on alternative wordings based on every-day or simplified language. This translation process may lead to new insights about the conceptual relations at stake (Prain, 2006).

Thus, GWPR instruction including genre knowledge, a lay audience, cognitive and metacognitive strategies for planning and revision necessitates students to reflect critically on their original understanding, stimulating their learning of new topic knowledge and insights. Additionally, GWPR supports students in understanding conceptual relations in texts of their academic discipline.

3.4 The process of writing-to-learn

In writing-to-learn research, the cognitive processes involved in learning while writing are scarcely investigated. Therefore, it is not clear how these processes can be observed, what they look like and what are differences between the processes of students who are learning while writing and those who are not. Two theories about the process of writing-to-learn have been proposed in the past decades and have been used as explanations for results found in empirical studies directed at newly learned topic knowledge and/or insights.

First, Bereiter and Scardamalia (1987) discriminated a writing process typical for experienced writers. They described it as a recursory process taking place in reviewing and revising the content, while reflecting on the goals. Their theory is that writers seek to reconcile contents and rhetorical goals, and therefore adjust their text several times on rhetorical and content aspects. They call this iterative process knowledge transformation, because pre-existing knowledge is transformed into new knowledge during writing. The second theory is the dual process theory proposed by Galbraith (2009). This theory distinguishes a knowledge retrieval process and a knowledge constituting process. Writers use their knowledge retrieval system, in which explicit knowledge is stored, for retrieving content taking into account rhetorical goals. While writing, they use their knowledge constituting system in order to make connections between concepts, some of which may be new connections the writer was not aware of previously (implicit knowledge). These new connections lead to new insights. The constitution of relations is a cyclic process in which writers alternately revise their text and refer to their knowledge constituting system. Both

theories state that learning by writing takes place in a cyclic process entailing reflection on content and rhetorical goals interactively (Klein, 2004).

Testing of these theories took place in various ways. Bereiter and Scardamalia (1987) provided planning instruction for students in grade 6 for writing an opinion essay and an exposition. They aimed at investigating how to instruct planning activities such that these lead to reflection. Experimental students were provided with cue cards for planning in a series of thirty-eight 50-minute lessons (divided over 19 weeks). Planning was modelled in several lessons by the researcher and students, and strategy instruction was provided for reconciliation of inconsistencies. Results were measured by means of a pre- and post-test, each requiring students to write an opinion essay and an exposition, without use of sources. Six students from the control group and six from the experimental group performed the planning of these four texts thinking aloud. The protocols were analyzed on the presence of reflective activities because reflection was considered as indicating knowledge transformation. The protocols of the experimental students showed an increase in reflective activities in the post-test, whereas (business as usual) control students showed a decrease. The difference between the two conditions on the post-test was significant, in favor of the experimental students. This is considered as evidence that experimental students progressed in transforming their knowledge during the planning of their texts in comparison to the control group.

Galbraith (2009) tested his dual process theory by comparing two types of writers (high self-monitors and low self-monitors) while writing essays, and measured writers' learning (that is, topic knowledge and insight) afterwards. High and low self-monitors differ in the way they operate when writing, well-considered (high self-monitors) or intuitively (low self-monitors). High self-monitors appeared to show larger effects of writing on learning than low self-monitors when making notes before writing, whereas low self-monitors showed larger effects on learning, when being assigned to write a text without any preparation. Galbraith considered the latter results as evidence for the existence of knowledge constituting. Because of their disposition, low self-monitors enter this process spontaneously without being directly aware of acquiring new knowledge.

Contrary to Bereiter & Scardamalia (1987), Galbraith (2009) did not measure the processes involved in writing-to-learn directly. However, in an experimental study, Baaijen & Galbraith (2018) used keystroke logging for measuring 78 university students' writing processes as described in the dual process theory. Change in insight was measured by comparing students' ratings of their insight just before and just after writing. Students' revision of text appeared to be related to increased insight.

Based on the theory of Bereiter & Scardamalia (1987) and an early version of Galbraith's (1992/1999) theory, Klein (2004) aimed to identify writing-to-learn processes by conducting an exploratory think aloud study with 56 university students. Students performed a science experiment and explained the outcome Then, they wrote a note (journal type) about how they arrived at their conclusion, while thinking aloud. Finally, they explained the outcome again. Change in insight was measured by comparing students' explanations before and after writing. A regression analysis provided evidence that students' reflection on goal setting, organizing and generating (planning) as well as reflection on reviewing and revising while writing promoted insight. Thus, whereas Bereiter and Scardamalia (1987) found evidence for reflection

in planning, and Baaijen and Galbraith (2018) in reviewing and revising. Klein (2004) found evidence in planning as well as in rereading and revising contents.

From the above, it follows that reflective processes mediating between rhetorical aspects and text contents, as proposed by Bereiter & Scardamalia (1987) as well as Galbraith (2009) are promising candidates for writing-to-learn. In the think aloud parts of the present studies, indicators of such reflective processes are more specifically defined. These indicators entail reflective activities (cognitive and metacognitive) performed during goal setting, organizing, generating, rereading and revising. Analysis of indicators of reflective processes involved in writing-to-learn can provide additional evidence for effects of GWPR on learning. By comparing verbalized writing processes of students from an experimental and a control group, we may find support for the expectation that GWPR stimulates reflective writing, which may lead to learning.

3.5 The present study

This study investigates whether GWPR instruction has effect on the learning of topic knowledge and insight in two widely differing disciplines in teacher education. Additionally, it investigates whether indicators of the writing-to-learn process can be identified by comparing think aloud protocols of experimental and control teacher candidates at the end of the intervention. Teacher education is an interesting context, because there are few empirical studies directed at the effects of writing-to-learn instruction in that context (Van Dijk et al., 2022a). In addition, it was a practical choice, because the first author is a teacher educator at the university involved and therefore well informed of organizational and personal issues that have to be taken into account.

The study was situated in biology (study 1) and mathematics (study 2) teacher education. This allowed us to compare results from two widely different disciplines regarding the role that writing plays in educational practice. In biology, writing is a relatively frequent activity (e.g., in lab reports), whereas writing in mathematics rarely occurs (Veel, 1999). The interventions were embedded in regular courses, as Miller et al. (2018) suggested. As Hunter & Tse (2013) suggested, the study was carried out in cooperation with the biology and mathematics teacher educators for organizing, composing, and embedding the intervention in their regular courses.

Apart from an experimental study directed at effects of GWPR, a think aloud part was included in each study. The addition of a process analysis to the effect study offers the opportunity to investigate whether the outcomes of the experiment and think aloud part complement each other. Our process analysis is intended to find empirical support for previous findings that different sorts of (cognitive and metacognitive) reflective activities indicate writing-to-learn (Baaijen & Galbraith, 2018; Bereiter & Scardamalia, 1987; Klein, 2004).

We formulated the following two research questions:

1. Does GWPR instruction lead to more topic knowledge and insight in the context of biology and mathematics teacher education, in comparison to a control group receiving business as usual lessons?

2. Does GWPR instruction lead to observable differences in processes of writing-to-learn between experimental and control teacher candidates?

3.6 Study 1 Writing-to-learn in biology teacher education

3.6.1 Materials and Methods

3.6.1.1 Participants

The study took place in a third-year biology teacher education course at a university. At the start, 53 teacher candidates participated. However, 15 teacher candidates did not perform the pre-test or post-test and were therefore excluded. Reasons for dropping out were illness, study break off, and moving. Consequently, 38 teacher candidates were left for analysis, 20 in the control group and 18 in the experimental condition. Table 3.1 presents age, gender, mother tongue and prior education. The latter was a high school degree for 28 students, whereas 10 students had followed higher education prior to entering biology teacher education.

For the analysis of the process of writing-to-learn, 12 teacher candidates were randomly selected from the 38 teacher candidates as participants of a think aloud study. Six teacher candidates belonged to the control group and six to the experimental condition. Table 3.2 shows characteristics of the 12 participants. One biology teacher educator was involved in the study. The teacher educator had 15 years of experience in higher education and was the regular teacher educator for both the control and the experimental group. Due to personal circumstances, he taught the control group in the first four lessons only. A biology colleague replaced him during the remaining four lessons in the control group.

Table 3.1 Characteristics of participants of biology teacher education

	Experimental group (N=18)	**Control group (N=20)**
Age	M: 27 (SD: 9.94)	M: 24 (SD: 3.73)
Gender	Female: 11	Female: 10
Mother tongue	Dutch: 16	Dutch: 19
Prior education	High school: 15	High school: 13

Table 3.2 Characteristics of selected participants of the think aloud part in biology teacher education

	Experimental group (N=6)	**Control group (N=6)**
Age	M: 27 (SD: 9.22)	M: 22 (SD: 1.51)
Gender	Female: 4	Female: 2
Mother tongue	Dutch: 5	Dutch: 6
Prior education	high school: 6	high school: 3

3.6.2 Design

We used a quasi-experimental, post-test-only design, comparing a control group with an experimental group. The dependent variables were topic knowledge and insight

into biology subject matter taught. Prior knowledge of biology and vocabulary knowledge were used as covariates. The control group received the regular biology lessons (business as usual), while the experimental group received the lessons including writing-to-learn tasks. The two groups received their lessons in two consecutive academic years, respectively 2011-2012 for the control group and 2012-2013 for the experimental group. In the third academic year of teacher education, only one class received lessons. This was the reason why we chose to conduct the study in two consecutive academic years. Otherwise the sample of participants would have been too small.

Observations of the control group lessons were organized in order to get acquainted with the objectives and structure of the regular course and for designing the experimental course, specifically how to embed writing-to-learn tasks in the lessons. For analyzing the process of writing-to-learn, a think aloud multiple case study was carried out with six experimental and six control teacher candidates. Utterances were coded and systematically analyzed in order to investigate differences between the (reflective) writing processes of experimental and control teacher candidates.

3.6.3 Treatment

The experiment took place in a course aimed at the history of biology, and scientists' contributions to biology. Observations of the lessons for the control group were carried out by the first author to describe the proceedings in the business as usual condition and to prepare replacing parts of the business as usual lessons with the writing tasks needed in the experimental condition. In preparation of each lesson in the control group, teacher candidates had to study one or two chapters from the textbook and answer open-ended questions about the contents. A pair of teacher candidates additionally prepared a presentation about next week's topic including hands-on activities for their classmates. The lessons consisted of discussing teacher candidates' questions about subject matter, and their answers on the open-ended questions. In the final part of each lesson, the two teacher candidates presented next week's topic.

In preparation of the intervention, the biology teacher educator and the first author cooperated in designing model texts, writing tasks as well as a teacher educator manual, combining their expertise in respectively biology and writing-to-learn. They discussed how to embed the writing tasks in the regular lessons and which part of the lessons would be left out. Furthermore, the first author proposed topics from the biology textbook and sources for the model texts and decided together with the teacher educator which would fit best for the writing tasks. The role of the teacher educator was to secure that the contents of the writing tasks and the aims of the biology course matched. Therefore, he evaluated the instructions included in the writing tasks, and whether the writing tasks focused on objectives of the biology lessons. Additionally, the teacher educator reacted to the suggestions in the teacher educator manual in order to decide whether it was sufficiently clear how he should present the conceptual relations involved.

The treatment in the experimental group is aimed at writing about the chapters studied, based on the principles of GWPR. Analysis of the assignments and texts used in the regular course showed that the most used conceptual relations were

comparison, sequence, and cause-effect. Thus, writing tasks in the genre 'explanation' would fit well with the objectives of the biology course (Rose, 2008). Therefore, we developed model texts and writing tasks each focusing on one of these three relations separately. Each of the three writing-to-learn tasks required teacher candidates to write an explanation directed at an audience of students in grade 9-10. This audience, with no knowledge about the topics, was important, because it required a thorough and simplified explanation from the teacher-candidates in their writing. This requirement added to the ecological validity of the writing tasks, because biology teacher education aims at teaching students in secondary education.

GWPR instruction entails that each writing task is preceded by an explanation of a model text. Model texts were derived from a biology textbook directed at the audience (grades 9-10) and rewritten to fit our needs. The topics of each model text and accessory writing task were related so that the teacher-candidates could use the model text for their writing. However, in order to avoid copy-paste strategies, the topics of the model text and the writing task were not identical (e.g., 'differences between regular medicine and alternative cure' for the model text, and 'a difference between Hippocrates' vision on medicine and the contemporary visions on alternative cure' for the writing task). Three model texts were composed, each containing examples of one conceptual relation. A conceptual relation was expressed in various ways (for instance the relation 'comparing' was represented by formulations such as 'bigger as...', 'compared to...', 'more important than...'). The purpose of this emphasis on different linguistic realizations of conceptual relations was to provide teacher-candidates with the instruments to consciously deal with the conceptual relations at stake in the chapters studied. The teacher candidates were expected to become aware of the relevance of these relations for their understanding of studied texts.

The experimental participants were explicitly instructed in each writing task to use the conceptual relation presented in the model text. In this way, the experimental participants were stimulated to transform their own thinking about the topic in order to accommodate the knowledge gap between themselves and their younger audience. We assume that such transformation is important for the teacher candidates to become aware of gaps in their explanation that need to be repaired in order to be understood by their readers (Prain, 2006).

Additionally, the writing tasks consisted of the following parts. The first was instruction on planning and entailed pairwise brainstorming. In their planning, experimental teacher candidates had to take the requirements into account posed by the given conceptual relation and the intended audience. The second part was writing a draft individually. The third part demanded pairs of teacher candidates to comment on each other's draft, while referring to the conceptual relation at stake and the audience, and, if necessary, to ask for clarification of each other's feedback. Finally, teacher-candidates had to revise their drafts individually by using their peers' feedback.

The biology teacher educator used the teacher educator manual (which was fabricated in cooperation with him) for presenting the model texts and the writing tasks to the teacher-candidates. It contained suggestions for explaining the conceptual relation to the teacher candidates, with the model text projected on a smart board and selected linguistic realizations of the conceptual relation marked. The teacher educator

also would ask teacher candidates to look for unmarked examples of the conceptual relation in the model text and discuss these.

3.6.4 Instruments

3.6.4.1 Prior knowledge tests

In consultation with the teacher educator, the researchers composed tests of prior knowledge about biology. The tests were based on the textbook that had been used during the first two academic years comprising eight major biological themes varying from plants to heredity (Campbell, 2011). The tests consisted of eight multiple choice (topic knowledge) and eight open-ended questions (insight), each referring to one of the themes.

The first author and the teacher educator coded teacher candidates' answers on the open-ended questions independently for a sample of eight cases. The inter-rater reliability was .81 (Pearson Correlation), which is considered acceptable. As can be expected, the homogeneity of the items measuring topic knowledge and insight was low, given that the items represented quite different areas of biological knowledge. Cronbach's alpha for the eight items on topic knowledge was .001 and for the eight items on insight .49. Despite the obvious heterogeneity of the tests, we decided to include both measures in our analysis. Cronbach's alpha provides an underestimation of test reliability (Boyle, 1991; Sijtsma, 2009; Taber, 2018). The tests for prior knowledge might still explain variance in our posttest measures. Therefore, we decided to include both measures as covariates in our analysis.

3.6.4.2 Vocabulary test

A vocabulary test of 30 items derived from the Dutch version of the Peabody picture vocabulary test (Dunn & Dunn, 2005) was composed. This Dutch version is based on frequencies per one million words. The words are ranked in 17 sets each aimed at a specific age group. For the selection of words, we used four sets (nr. 14 -17) aimed at ages above 18. We selected 30 words of which the expected proficiency was between p= .36 and p= .86. Cronbach's alpha of the vocabulary test was .83, which is considered acceptable.

3.6.4.3 Topic knowledge and insight in the post-test

The post-test served as final examination of the biology course and was designed in consultation with the biology teacher educator. It consisted of 30 multiple-choice questions (as a measure for topic knowledge), and nine open-ended questions (as a measure for insight). The multiple-choice questions as well as the open-ended questions covered the six periods in history distinguished in teacher candidates' textbook. Nine of the multiple-choice items correlated negatively with the remaining items, and therefore were removed. Cronbach's alpha of the remaining 21 items was .53. One open-ended question was excluded from the post-test, because of an unclear formulation. The researcher and the biology teacher educator coded teacher candidates' answers on the remaining open-ended questions of the post-test independently for a sample of eight cases. The inter-rater reliability was .88 (Pearson Correlation). Cronbach's alpha was .76 for eight open-ended questions. Although the

items for topic knowledge had a rather low reliability, we decided to include both parts of the test in our analysis.

3.6.4.4 Post-test writing task

A final writing task was assigned to all teacher candidates. This task required them to read two chapters (one mandatory and one free choice) of Darwin's 'The origin of species', to track down Darwin's statements, to look for observations these were derived from, and to describe them. The writing task was a regular instrument used by the teacher educator for testing teacher candidates' insight into Darwin's work.

Although the format of this writing task was quite different from the experimental writing-to-learn tasks used during the lessons, we used it in the think aloud study for examining the process of writing-to-learn. 'The origin of species' was not studied during the course. Therefore, this writing task barely measured understanding of course content. However, the task required teacher candidates' acquiring knowledge of subject matter by writing. Therefore, we considered it an appropriate measure for examining the process of writing-to-learn.

3.6.5 Procedure

Table 3.3 presents the lesson structure in the control group and in the experimental group. The business-as-usual lesson structure was as follows: posing questions about subject matter studied, discussion of teacher candidates' answers on the open-ended questions, and finally a presentation prepared by a pair of teacher candidates. The experiment comprised nine lessons, one lesson each week. The first lesson was used for administration of the prior knowledge and vocabulary tests; lessons 2-7 were regular lessons; lesson 8 was used for preparation of the post-test, the last lesson for the post-test and the final writing task.

In the experimental condition, the writing-to-learn tasks were embedded in lessons 2-7 as follows. Each writing task was divided into two parts in such a way that teacher candidates wrote a draft in one lesson and revised their text in the next (see experimental condition in Table 3.3). For securing treatment fidelity, the first author observed whether the teacher educator carried out the lessons as intended and as described in the teacher educator manual in all experimental lessons. No deviations were encountered. For keeping time on task equal in the control and experimental conditions, the part in which control teacher candidates were allowed to pose questions about subject matter was left out in the experimental condition and replaced by one part of the writing tasks.

In the academic year (2012-2013) after the control teacher candidates participated in the biology course, a change in timetabling of the teacher education institute resulted in longer lesson duration for the experimental lessons (150 instead of 100 minutes). A 15 minutes pause was inserted, but still the duration of experimental lessons was longer (some 35 minutes) than the control lessons. Unfortunately, we therefore cannot exclude that this difference in time-on-task has influenced the results of the experiment.

Table 3.3 Lesson structure in control and experimental condition of biology teacher education

Control group (N=20)	Experimental group (N=18)	
Business as usual (lessons 2-7)	**First draft (lessons 2, 4, 6)**	**Revision (lessons 3, 5, 7)**
• Posing questions about subject matter studied • Class discussion about students' answers on open-ended questions • Presentation of next week's topic by a small group of students	• Class discussion about students' answers on open- ended questions • Presentation of next week's topic by a small group of students • ***Writing-to-learn task, part 1***: **explanation of model text, planning and writing a first draft (on laptops)**	• ***Writing-to-learn task, part 2*: feedback and writing final text** • Class discussion about students' answers on open-ended questions • Presentation of next week's topic by a small group of students

Bold text indicates replacement of elements in the control condition by writing-to-learn tasks

Think aloud procedures can inform about the cognitive processes involved in task execution of several kinds (Ericsson and Simon, 1993; Van Someren, Barnard & Sandberg, 1994). For the think aloud experiment, the sample of 12 teacher candidates carried out their post-test writing task individually in an empty classroom, in the presence of the first author. She told them that she was interested in how they addressed the writing task, and for this reason, she asked them to think aloud while writing. She provided an instruction including a video clip of a student thinking aloud while writing a paper. Teacher candidates wrote their texts on a computer while using Darwin's 'The origin of species' and were free to use their self-made summaries of the chapters as well. They had to execute their task in one hour maximum. When teacher candidates kept silent for 10 seconds, the researcher encouraged them to keep thinking aloud, and used prompts such as: say aloud what you are thinking. The sessions were video recorded.

3.6.5.1 Coding teacher candidates' transcribed utterances

Teacher candidates' verbalizations were transcribed and represented in protocols as separate utterances in case of verbal behavior, and separate instances in case of non-verbal behavior (for instance: 'sighing'). An utterance was defined as a phrase containing a meaningful element of information, (Pander Maat, 1994).

The codes were based on Hayes and Flower's (1980) writing model. In total, 24 codes were used to describe teacher candidates' writing and thinking processes. For instance, the utterance 'all claims and observations have to be selected' was coded as 'thinking about task approach'. In addition, interruptions and utterances not focusing on the writing task were coded and attributed to two categories: interruptions and other remarks. Finally, the resulting coding scheme comprised 29 verbal and non-verbal activities (see Appendix A). By means of this coding scheme, the first author and a trained research-assistant coded the utterances, one code per utterance or per instance. For determining inter-rater reliability, two protocols (one from the experimental and one from the control group) were coded by two independent raters.

There was agreement for 84% of all utterances, a fair amount for our purposes. Differences in coding were resolved after discussion.

Reflective activities such as reviewing, revising, goal setting, organizing and generating contents can be regarded as indicators of writing-to-learn processes, according to previous studies (Bereiter & Scardamalia, 1987; Baaijen & Galbraith, 2018; Klein 2004). The following specific codes from our list can be regarded as indicators for these reflective activities: 1) using knowledge about audience, 2) thinking about content selection, 3) thinking about formulating, 4) revising while formulating, 5) revising after finishing an utterance, 6) rereading own text, 7) rethinking task approach.

3.6.6 Data Analysis

3.6.6.1 Prior knowledge, Vocabulary Test and Post-test

Two ANOVA's were used for comparing prior knowledge (insight) and vocabulary in the two conditions. By means of two ANCOVA's, teacher candidates' post-test scores in the two conditions were compared. The two dependent variables were the sums of teacher candidates' scores on topic knowledge and insight. The sums of teacher candidates' scores on prior knowledge (insight) and on vocabulary were used as covariates. In all statistical tests, alpha level was .05.

3.6.6.2 Process analysis

For determining whether utterances that were considered indicators for writing-to-learn occurred in a larger frequency in the experimental than in the control group, the means of each code per condition were computed. Subsequently, the ratio of the mean frequency of each code to the total number of utterances in a condition was computed. Finally, effect-sizes (Cohen's d) were used to evaluate differences between the two groups in proportioned mean frequencies. As Cohen (1988) suggests, effect sizes below .20 were considered as no effect, the range between .20 and .50 was considered a small effect, the range between .50 and .80 a medium effect and ≥ .80 a large effect.

3.6.7 Results

Table 3.4 presents the means and standard deviations for the four variables involved, including the pre-tests for prior insight, prior topic knowledge, and vocabulary, and the post-tests for insight and topic knowledge.

Table 3.4 Means and standard deviations for prior insight and topic knowledge, vocabulary knowledge and post-test scores on insight and topic knowledge in biology teacher education

Variables	Experimental group N=18) Mean (SD)	Control group (N=20) Mean (SD)
Prior insight[1)]	3.81 (1.66)	4.50 (1.61)
Prior topic knowledge[2)]	3.83 (1.15)	4.35 (1.27)
Pre-test vocabulary[3)]	18.11 (6.40)	19.70 (4.61)
Post-test insight[4)]	5.28 (1.66)	3.40 (1.82)
Post-test topic knowledge[5)]	13.33 (2.64)	10.85 (2.41)

[1)] theoretical maximum score: 15.5 [2)] theoretical maximum score: 8 [3)] theoretical maximum score: 30 [4)] theoretical maximum score: 13.5 [5)] theoretical maximum score: 21

3.6.7.1 Prior knowledge

An ANOVA showed that experimental and control students' scores on prior insight did not differ significantly, $F(1,36) = 1.712$, $p = .199$, nor did the two groups differ significantly on prior topic knowledge, $F(1,36) = .716$, $p = .198$.

3.6.7.2 Vocabulary

An ANOVA showed no significant differences between the experimental and control teacher candidates' scores on pre-test vocabulary: $F(1,36) = 0.783$, $p = .382$.

3.6.7.3 Post-tests

Two ANCOVA's were conducted to compare students' scores on post-test insight and post-test topic knowledge in the experimental and control condition. The scores on prior insight, prior topic knowledge and on vocabulary served as covariates. The ANCOVA's showed that the covariate prior insight significantly predicted the scores on post-test insight, $F(1,36) = 4.479$, $p = .042$, partial $\eta^2 = .120$ (medium), as well as on post-test topic knowledge, $F(1,36) = 5.997$, $p = .020$, partial $\eta^2 = .154$ (large). The covariate prior topic knowledge did not predict differences in post-test insight, $F(1,36) = .218$, $p = .643$, nor in topic knowledge, $F(1,36) = 3.393$, $p = .074$. The covariate vocabulary did not predict differences in insight $F(1,36) = .073$, $p = .789$, nor on topic knowledge: $F(1,36) = .774$, $p = .385$. Therefore, in the final analysis, prior topic knowledge and vocabulary were not included as covariates, while prior insight was.

This analysis shows that the scores of experimental and control students on the post-tests of insight and topic knowledge differed significantly, $F(1,36) = 15.023$, $p = .00$, partial $\eta^2 = .30$ (large) for insight, and $F(1,36) = 13.43$, $p = .001$, partial $\eta^2 = .28$ (large) for topic knowledge. It appears that the experimental students profited from the intervention consisting of writing-to-learn with GWPR and outperformed the control students in the business-as-usual condition.

3.6.7.4 Characteristics of writing-to-learn

In Table 3.5, the first column comprises seven (of 29) codes indicating teacher candidates' verbal behavior. We predicted that these codes indicate reflection and therefore writing-to-learn (see Appendix A for an overview of all 29 codes).

The remaining columns show the proportioned mean frequencies and standard deviations of codes for the experimental and control condition and effect size (Cohen's *d*) as an estimation of the magnitude of the difference between the two groups.

Table 3.5 Proportioned means, standard deviations, and effect sizes of indicators of writing-to-learn in experimental (N=6) and control group (N=6), in biology teacher education

	Experimental	Control	Effect size
Codes	**Mean (SD)**	**Mean (SD)**	**Cohen's *d***
PLANNING: generating			
Using knowledge about audience	.00 (.00)	.00 (.00)	.00
PLANNING: selecting			
Thinking about content selection	.03 (.04)	.04 (.04)	-.27
FORMULATING			
Thinking about formulating	.02 (.02)	.02 (.03)	.00
Revising while formulating	.03 (.03)	.02 (.01)	.49
Revising after finishing an utterance	.03 (.04)	.01 (.01)	.75
MONITORING			
Rereading own text	.08 (.09)	.03 (.03)	.82
Rethinking task approach	.01 (.01)	.01 (.01)	.00

We expected to find differences between the (proportioned) mean frequencies, in favor of the experimental condition. Table 3.5 shows that differences in three activities are relatively large and in the expected direction: revising while formulating *(*e.g., *he needs a tree. . .an apple tree to be able to grow*), *d*= .49 (small effect), revising after finishing an utterance (e.g., *this happens two ... this is... no no, this has two reasons), d*= .75 (medium effect), and rereading own text, *d*= .82 (large effect).

The other four hypothesized activities did not show larger frequencies for experimental teacher candidates. The first, using knowledge about audience, was not applied by the teacher candidates in both conditions. The second, thinking about content selection, was performed more often by control teacher candidates (*d* = - .27). The third, thinking about formulating, was equally frequent in both conditions and so was the fourth activity, rethinking task approach. Thus, it appeared that teacher candidates who had received GWPR-instruction, showed three out of seven of the hypothesized activities more often than the control group.

3.7 Study 2: Writing-to-learn in mathematics teacher education

3.7.1 Materials and methods

3.7.1.1 Participants

This study took place in a third-year mathematics teacher education course at a university. The control group started with 51 teacher candidates and the experimental group with 38 teacher candidates. However, 27 teacher candidates were excluded. Reasons were enrolment for a resit of the final test only (because of failure in the year before), or attendance of just a few lectures (teacher candidates' presence was not obligatory). In the analyses, 62 teacher candidates were included, 36 in the control group and 26 in the experimental condition. Table 3.6 presents age, gender, mother tongue and prior education of the participants. Part of the participants (29) possessed a high school degree only, but most of them (33) had followed higher education prior to their enrolment in mathematics teacher education.

For answering the second research question concerning the process of writing-to-learn, 15 mathematics teacher candidates were randomly selected from the sample of 62 teacher candidates as participants in the think aloud study. Seven teacher candidates belonged to the control group and eight to the experimental condition. Table 3.7 presents their characteristics. One mathematics teacher educator was involved in the study. He had nine years' experience in teacher education and taught both conditions.

Table 3.6 Characteristics of the participants of mathematics teacher education

	Experimental group (N=26)	**Control group (N=36)**
Age	M: 26 (SD: 9.64)	M: 29 (SD: 11.25)
Gender	Female: 14	Female: 18
Mother tongue	Dutch: 22	Dutch: 33
Prior education	High school: 10	High school 19

Table 3.7 Characteristics of the selected participants for the think aloud study in mathematics teacher education

	Experimental group (N=8)	**Control group (N=7)**
Age	M: 20 (SD: 1.06)	M: 21 (SD: 2.27)
Gender	Female: 3	Female: 3
Mother tongue	Dutch: 8	Dutch: 6
Prior education	Highschool: 3	Highschool: 4

3.7.2 Design

The design of this study was the same as Study 1. The lessons and writing tasks were about topics from mathematics (sequences and limits). The experiment took place in two consecutive academic years and started one year later than Study 1: the control group in 2012-2013 and the experimental condition in 2013-2014. In the third academic year of mathematics teacher education only one class received education. This was the reason why we chose to conduct the study in two consecutive academic years.

3.7.3 Treatment

The experiment took place in a course aimed at insight into linking rows and recurrent relations, the use of web graphs for computing these relations, and computing limits. As in study 1, observations were carried out by the first author in order to describe the proceedings in the business as usual condition (the control group) and to plan adaptations for the writing tasks in the experimental condition. In the control group, teacher candidates studied sections of a chapter from their textbook and completed a number of additional assignments as their weekly homework. In the lessons, the teacher educator discussed the tasks performed at home using a whiteboard. Furthermore, new topics were introduced.

The intervention in the experimental group aimed at elaborating on the homework and entailed the implementation of the principles of GWPR. On the basis of the observations in the control group, it was concluded that the genre explanation would fit well with the aims of the lessons. Analysis of the textbooks and assignments revealed that the conceptual relations 'condition', 'definition' and 'sequence' belonged to the dominant mathematical reasoning and explaining. Therefore, it was decided to make teacher candidates sensitive to these conceptual relations in the experimental lessons.

The researcher and the teacher cooperated in developing two writing tasks, two model texts and a teacher educator manual. The writing tasks required teacher candidates to write an explanation directed at an audience (grade 10 students) that they taught in their apprenticeship. Therefore, the model texts were based on textbooks directed at grade 10. The topics of the model texts were closely related to the topics of the accessory writing tasks (for instance, 'limits' for the model text and 'computing limits' for the writing task). In the model texts, the conceptual relations were presented in various wordings (for instance for the relation sequencing: first…, thereafter…). As usual in texts about mathematics (Veel, 1999), the model texts comprised graphical representations, such as a table or formulas in addition to the text. The way instruction in writing-to-learn was applied, was the same as in Study 1 (see section 3.6.3).

3.7.4 Instruments

3.7.4.1 Prior knowledge

The researchers composed the tests of prior knowledge in consultation with the teacher educator. The tests were based on textbooks from Bos et al. (2007), which had been used during the first academic year in four courses directed at the mathematical field 'analysis' providing prior knowledge for the course central in the present study. Five themes, varying from function theory to differential equation, had been taught. The tests consisted of eight multiple choice (topic knowledge) and six open-ended questions (insight), referring to the themes.

The first author evaluated teacher candidates' answers on the open-ended questions in consultation with the teacher educator. Inter-rater reliability was not computed, because of the unambiguity of the answers, consisting of definitions of concepts required for three open-ended questions and solutions for the remaining mathematical tasks.

As can be expected, the homogeneity of the items measuring topic knowledge and insight was low, given that the items represented various themes from the field of analytics analytics that is part of mathematics. For the eight items on topic knowledge Cronbach's alpha indicated a large heterogeneity (–.118). The six items on insight were more homogeneous, as indicated by a Cronbach's alpha of .56. Despite the heterogeneity of the tests, we decided to include both measures in our analysis. As mentioned in study 1, Cronbach's alpha provides an underestimation of test reliability (Boyle, 1991; Sijtsma, 2009; Taber, 2018). Therefore, the tests for prior knowledge might still explain variance in our posttest measures.

3.7.4.2 Vocabulary

The vocabulary test consisting of 30 items described in Study 1, was used in Study 2 as well. One item appeared to correlate negatively with the rest. Therefore, this was removed. Cronbach's alpha of the test consisting of 29 items was .86, which is considered acceptable

3.7.4.3 Topic knowledge and insight in the post-test

The post-test consisted of four multiple-choice items and seven open-ended questions. The homogeneity of the multiple-choice test was .15 (Cronbach's alpha). We explain this low homogeneity by the small number of items in the test. Because of the low homogeneity, the results were not included in the analyses.

The seven open-ended questions were mathematical tasks and belonged to the usual final test. We left its evaluation with the teacher's expertise. Cronbach's alpha for the open-ended questions was .81, which is considered acceptable.

3.7.4.4 Post-test writing task

For examining the process of writing-to-learn in the think aloud study, a writing task was added to one of the mathematical tasks in the post-test. First, students were required to carry out one mathematical task.The subsequent writing task entailed the explanation (in language) of the way students had calculated their outcome of the mathematical task by focusing on the theorem they were instructed to apply. They had to write their explanation for an audience of grade 10 students.

3.7.5 Procedure

The lesson structure for the control group was as follows: the teacher educator discussed teacher candidates' questions about their homework. Thereafter, he lectured teacher candidates about new theory, while representing this in mathematical symbols and schemes on a whiteboard. The course lasted nine weeks: eight lessons lasting 100 minutes each and a final examination in the ninth week.The writing-to-learn tasks for the experimental group were embedded in four of eight lessons as follows. For each writing task students wrote a draft in one lesson and a revised text in the next (see lessons 4 -7 in Table 3.8). For securing treatment fidelity, the first author observed whether the teacher educator carried out the lessons as intended and as described in the teacher educator manual in all experimental lessons. No deviations were encountered.

Time on task remained equal for both conditions by replacing parts of the discussion of students' questions and the introduction of new theory by the writing tasks in the experimental lessons. The prior knowledge and vocabulary tests were administered in the first lesson for both conditions. The second, third and eighth lesson were identical for both conditions. The post-test including the post-test writing task was administered in the ninth week of the course in both conditions.

Table 3.8 Lesson structure in control and experimental condition in mathematics teacher education[1]

Control group	**Experimental group**	**Experimental group**	
Lessons 2 – 7	**Lessons 2, 3**	**Lessons 4, 6**	**Lessons 5, 7**
Business as usual		**First draft**	**Final text**
• Discussion of students' questions about subject matter studied.		• Discussion of students' questions about subject matter studied.	• Discussion of students' questions about subject matter studied.
• Lecture about new theory		• **Writing-to-learn[2] task, part 1: explanation of model text, planning and writing a first draft (by hand)**	• **Writing-to-learn task, part 2: feedback and writing final draft**
		• Lecture about new theory	• Lecture about new theory

[1]) Lessons 1, 2, 3, 8 and 9 were identical for both groups. [2]) Bold text indicates replacement of elements in the control condition by writing-to-learn tasks.

For the think aloud experiment, the sample of 15 students completed the writing task thinking aloud. Therefore, they completed only the mathematics task during the post-test session and performed the additional writing task immediately after the post-test. Students executed the writing task individually in an empty classroom, in the presence of the first author. The researcher's behavior was as described in Study 1. Students wrote their texts on a computer thinking aloud, while having their computation of the mathematics task at hand. They had to perform the writing task in maximally half an hour. The sessions were video recorded.

3.7.5.1 Coding teacher candidates' transcribed utterances

In Study 1, we explained how the coding scheme was composed. For Study 2, this scheme consisting of 29 codes was used as well (see Appendix B). The first author and a trained research-assistant coded utterances (verbal behavior) and instances (nonverbal). Two complete protocols (one for each condition) were coded independently by the two raters. There was agreement for 85% of all utterances and instances. We consider this a sufficient reliability of coding. Differences in coding were resolved after discussion. From these codes the same (seven) codes as in Study

1 were selected as indicating the process of writing-to-learn (cognitive and metacognitive processes).

3.7.6 Data analysis

3.7.6.1 Prior knowledge, vocabulary and post-test

To test for equivalency of groups, three ANOVA's were used comparing students' prior topic knowledge, prior insight and vocabulary in both conditions. To test for differences between students' insight, we compared the post-test scores (open-ended questions) of the two conditions by conducting an ANCOVA. Prior topic knowledge, prior insight and vocabulary were used as covariates.

3.7.6.2 Process analysis

The analysis of the process of writing-to-learn was performed as in Study 1 (see section 3.6.6.2).

3.7.7 Results

Table 3.9 shows teacher candidates' mean scores and standard deviations on the pre-tests for prior insight, prior topic knowledge and vocabulary, and on the post-test for insight.

Table 3.9 Means, standard deviations for prior insight and topic knowledge, vocabulary knowledge and post-test scores on insight in mathematics teacher education

Variables	Experimental group (N=26) Mean (SD)	Control group (N=36) Mean (SD)
[1)] Prior insight	4.27 (2.78)	3.85 (2.37)
[2)] Prior topic knowledge	3.96 (1.18)	4.17 (1.30)
[3)]) Pre-test vocabulary*	14.59 (4.57)	17.11 (6.97)
[4)]) Post-test insight	28.35 (11.42)	21.22 (11.50)

[1)] theoretical maximum score: 23 [2)] theoretical maximum score: 8 *four students did not take the test
[3)] theoretical maximum score: 29 [4)] theoretical maximum score: 49

3.7.7.1 Prior Knowledge

The differences between students' prior insight in the two conditions were not significant $F(1,60) = .414$, $p = .522$. The differences between students' prior topic knowledge in the two conditions were also not significant $F(1,60) = .406$, $p = .527$.

3.7.7.2 Vocabulary

There were no significant differences between teacher candidates' vocabulary in the two conditions $F(1,56) = 2.27$, $p = .137$.

3.7.7.3 Post-test

The ANCOVA showed that the covariates prior insight, prior topic knowledge and vocabulary did not significantly predict the scores on insight in the post-test, respectively $F(1,56) = 1.829$, $p = .182$, $F(1,56) = .147$, $p = .703$ and $F(1,56) = 3.216$, $p = .079$. Therefore, in the final analysis the covariates were not included. An

ANOVA showed that the scores on the post-test insight of control and experimental students differed significantly, $F(1,60) = 5.829$, $p = .019$, partial $\eta^2 = .089$ (medium effect). Experimental students outperformed control students in post-test insight scores.

3.7.7.4 The process of writing-to-learn

Table 3.10 shows proportioned frequencies of the seven selected indicators of writing-to-learn (reflective activities) for both conditions. The results for all 29 coded activities of the think aloud study can be found in Appendix B. Effect sizes are shown for estimating the magnitude of differences in proportioned frequencies between the two conditions.

Table 3.10 Proportioned means, standard deviations, and effect sizes of indicators of writing-to-learn in experimental and control group of mathematics teacher education

	Experimental (N=8)	Control (N=7)	Effect size
Codes	**Mean (SD)**	**Mean (SD)**	**Cohen's *d***
PLANNING: generating			
Using knowledge about audience	.04 (.03)	.02 (.02)	0.83
PLANNING: selecting			
Thinking about content selection	.03 (.05)	.01 (.01)	0.55
FORMULATING			
Thinking about formulating	.03 (.02)	.04 (.07)	- 0.25
Revising while formulating	03 (.04)	.01 (.02)	0.38
Revising after finishing an utterance	.02 (.02)	.02 (.02)	0.00
MONITORING			
Rereading own text	.12 (.09)	.04 (.06)	1.27
Rethinking task approach	.01 (.01)	.01 (.02)	0.00

It appears that there are differences in the expected direction for four activities: using knowledge about audience (e.g., the utterance: *for students I would explain it by means of a stable population*), $d = .83$ (large effect), thinking about content selection, $d = .55$ (medium effect), revising while formulating, $d = .38$ (small effect) and rereading own text, $d = 1.27$ (large effect). The remaining three hypothesized indicators of writing-to-learn did not show differences between the conditions in favor of the experimental teacher candidates. The activity thinking about formulating was slightly more frequent for the control teacher candidates, $d = -.25$. The second formulating activity revising after finishing an utterance was performed just as often by both conditions, $d = .00$. The monitoring activity rethinking task approach (e.g., *I have to do it in another way...)* was equally frequent in both conditions as well, $d = .00$. Thus, in these three cases, the hypothesis was not confirmed.

3.8 Discussion

3.8.1 Conclusions

We expected that GWPR instruction comprising genre knowledge, planning and revision, and writing for a lay audience, creates favorable conditions for the process of writing-to-learn. These entail teacher candidates' awareness of characteristic conceptual relations in the genre at stake, teacher candidates' reflection during generating and organizing text contents and during rereading and revising the formulation of the conceptual relations, aiming at comprehensibility of the text to a lay audience.

The results of both studies showed that GWPR instruction leads to effects on insight and topic knowledge in biology teacher education (with large effect sizes), and on insight in mathematics teacher education (large effect size). Because the post-test for topic knowledge on mathematics was not reliable, results for topic knowledge could not be included in the analysis. The aim of the two think aloud studies was to identify indicators of the writing-to-learn process. The hypothesis was that reflective activities (cognitive and metacognitive) performed during organizing and generating contents and reviewing and revising are indicative for the process of writing-to-learn. Therefore, it was expected that experimental teacher candidates performed reflective activities more often than control teacher candidates.

For reader's convenience, Table 3.11 presents the outcome of the analysis of the process of writing-to-learn for both studies. Effect sizes are presented for estimation of the magnitude of differences found between experimental and control teacher candidates. In biology teacher education, some evidence for the hypothesis was found. Experimental teacher candidates executed three of seven reflective activities more often than control teacher candidates. In mathematics teacher education, four reflective activities were more often carried out by teacher candidates who had received GWPR instruction than by control teacher candidates. Two of these (revising while formulating and rereading own text) were similar to two indicators of writing-to-learn in biology teacher education.

This evidence partly confirmed our view on the role of reflective activities in writing-to-learn. Differences found between the two studies can be explained by differences between the two writing tasks in biology and mathematics. The outcome of the think aloud studies is complementary to the results found for topic knowledge and insight in the two studies. It provides evidence that GWPR instruction incites the process of writing-to-learn by teacher candidates' reflection on conceptual relations and comprehensibility to the intended audience, leading to more insight and topic knowledge in biology teacher education and insight in mathematics teacher education.

Table 3.11 Overview of the results of the process analyses in Study 1 and Study 2

Indicators of writing-to-learn	Study 1: Biology Cohen's d	Study 2: Mathematics Cohen's d
Planning		
Using knowledge about audience	Not used	Large
Thinking about content selection	Small negative effect	Medium
Formulating		
Thinking about formulating	No effect	Small negative effect
Revising while formulating	Small	Small
Revising after finishing an utterance	Medium	No effect
Monitoring		
Rereading own text	Large	Large
Rethinking task approach	No effect	No effect

Not used: not used by both conditions

3.8.2 Effects of GWPR instruction

In the two studies presented, GWPR instruction appeared to enhance learning in biology and mathematics teacher education. The strength of GWPR instruction may be that it initiates a coherent writing process by focusing teacher candidates' attention repeatedly on genre-specific formulation of conceptual relations and comprehensibility to a lay audience. Therefore, the instructed activities link up with each other. We assume that this explains why the process resulted in new topic knowledge and insight in biology and insight in mathematics.

To recapitulate, genre writing with planning and revising (GWPR) starts with discussing a model text. This text has the characteristics of the genre (e.g., exposition), (lay) audience and topic the teacher candidates are supposed to write about, but it is designed in such way that copying contents is prohibited. In addition, in the model text examples of formulations of a conceptual relation are highlighted and reflected upon in class. Subsequently, teacher candidates are stimulated to reflect on their planning activities in small groups by discussing selection of content elements and how to clarify the relationships between these elements in a way that is comprehensible to the lay audience, taking advantage of examples of linguistic realizations in the model text. In the context of teacher education, we could make use of the fact that it is directed to the teaching of younger teacher candidates. This provided a good occasion for an ecologically valid audience for writing. After teacher candidates have written a draft on their own, they are stimulated to provide feedback on each other's drafts in pairs or small groups with specific attention to the comprehensibility of the conceptual relations to the audience. Finally, teacher candidates use this feedback, for revising.

3.8.3 Teacher candidates' and teacher educators' evaluation of writing-to-learn

Apart from learning effects of GWPR, it is also relevant to find out how teacher candidates and teacher educators value our approach to writing-to-learn. Therefore, we asked how they evaluated the use of GWPR instruction in their classes. Do teacher candidates believe that they benefit from writing-to-learn? Do teachers consider writing-to-learn a useful addition to their teaching repertoire? For answering these

questions, teacher candidates' and teacher educators' views were explored in individual interviews (three biology teacher candidates, three mathematics teacher candidates and two teacher educators) that took place after finishing the experiments. In addition, a questionnaire was administered to experimental teacher candidates (18 biology and 26 mathematics teacher candidates).

All biology teacher candidates considered writing supportive for learning and mentioned that they acquired new topic knowledge and insight. They were positive about using model texts stating that these made clear which type of text the teacher educator expected. The biology teacher educator appreciated the teacher educator manual for familiarizing him with GWPR instruction, but still felt insecure about explaining the conceptual relations in the model texts. However, his intention was to continue using writing-to-learn in his lessons.

Mathematics teacher candidates' reaction on GWPR instruction was mixed. The three interviewed mathematics teacher candidates experienced the writing activities as useful and meaningful. Two teacher candidates valued peer feedback, because this made them realize that their texts were not understandable to an audience yet. However, in their evaluations quite a number (15 of 26) of mathematics teacher candidates appeared to be not convinced of the usefulness of writing-to-learn assignments. One mathematics student was afraid not to be able to write a text as long as the model text.

Their teacher educator did not feel comfortable with teaching writing-to-learn. He wondered if the course 'Rows and limits' was suitable for using writing-to-learn tasks and suggested that a course requiring teacher candidates to write mathematical proofs might offer better opportunities for writing-to-learn.

These reactions reflect differences between the two disciplines. The disciplines involved in the experiment differ largely regarding the role of writing. In biology teacher education, teacher candidates are used to write about subject matter, whereas mathematics teacher candidates usually do not write texts in math classes. In class, they are not challenged to express their knowledge of subject matter in their own words (Veel, 1999; Skemp, 1987), which is an important element of writing-to-learn. Therefore, it is remarkable that mathematics teacher candidates showed positive effects on learning. Using writing-to-learn tasks in mathematics teacher education may entail a much larger pedagogical change than in biology teacher education (Graham et al., 2020).

3.8.4 The process of writing-to-learn

The assumption behind the think aloud studies was that GWPR instruction stimulates reflection on content and rhetorical goals (Bereiter & Scardamalia, 1987). Seven reflective activities were identified as indicators of this process. The biology as well as the mathematics experimental teacher candidates performed two of these activities, namely 'revising while formulating' (small effect sizes) and 'rereading own text' (large effect sizes), more frequently than control teacher candidates. This is in accordance with the discussed theories, which stress the importance of rereading and revising for learning by writing (Bereiter & Scardamalia, 1987; Galbraith, 2009). Additionally, these similarities between the two types of teacher education provide support for Klein (2004) who showed evidence for rereading and revising.

Differences between the two types of teacher education can largely be explained by differences between the writing tasks and teacher candidates' familiarity with writing. The mathematics post-test writing task was similar to the writing-to-learn tasks teacher candidates performed in the lessons (assignment of genre and a lay audience), but the biology post-test writing task was not (no assignment of genre and no lay audience). In the first place, the biology writing task contained criteria for selecting contents ('look for Darwin's observations and his accessory explanations'). Therefore, it is understandable that experimental biology teacher candidates did not differ from control teacher candidates on the indicator 'thinking about content selection', whereas the experimental mathematics teacher candidates did. In addition, biology teacher candidates were not instructed to write for an audience, contrary to mathematics teacher candidates. Therefore, it is not surprising that biology experimental teacher candidates did not differ from control teacher candidates on the indicator 'using knowledge about audience'.

Another difference concerns experimental mathematics teacher candidates not showing the indicator 'revising after writing an utterance' more frequently than control teacher candidates, whereas experimental biology teacher candidates did. This can be explained by the previously mentioned unfamiliarity of mathematics teacher candidates with writing. Therefore, they might have been hesitant to revise their text, a phenomenon that is often encountered in inexperienced writers (Beal, 1990; Bereiter & Scardamalia, 1987; Van Gelderen, 1997).

The indicator 'thinking about formulating' did not show larger frequencies for biology nor for mathematics experimental teacher candidates in comparison to control groups. For biology teacher candidates, this can be explained by the writing task as well. Both conditions did not need 'thinking about formulating', because they disposed of Darwin's formulation. Therefore, the conditions may have acted in the same way. Mathematics teacher candidates did not like writing, as previously described. Therefore, they probably were not inclined to spend much effort (and reflection) on their formulation processes. In biology as well as mathematics teacher education, experimental teacher candidates did not differ from control teacher candidates on the indicator 'rethinking task approach'. Possibly, teacher candidates in both studies did not consider reflecting on their task approach, because both writing tasks provided enough structure for them to follow. In that case, they did not see a reason to critically evaluate their approach.

3.9 Suggestions for future research

In the two studies reported, GWPR instruction facilitated academic learning in two largely differing disciplines in higher education. Additionally, teacher candidates valued learning by writing, albeit more in biology than in mathematics. Therefore, we believe that future research into the effects of GWPR instruction on topic knowledge and insight is worthwhile in order to determine the generalizability and stability of these findings in other educational contexts. An important issue to concern is the role of teacher attitudes towards writing-to-learn (see Pedagogical Implications).

This applies to both higher and secondary education in different disciplines. Relatively, much research on writing-to-learn has already been conducted in secondary education (Miller et al., 2018; Graham et al., 2020). In that context,

research into effects of GWPR instruction however is of value, since it is still unknown what type of instruction for writing-to-learn is more effective than others (Graham et al. o.c.). GWPR instruction offers much support in understanding and producing conceptual relations in writing, which may be beneficial for learning processes of students in secondary school.

In the present studies in higher education, the numbers of participants (including teacher educators) were quite small. For providing stronger evidence, we recommend research on larger samples. This may be realized in higher education courses teaching larger numbers of teacher candidates, for example in their first year of study. Alternatively, a cooperation of several universities teaching the same course may be considered. A requirement is that teachers involved are motivated for working in a team consisting of researchers and teachers from various universities for cocreating materials and lessons and aligning their assessments of learning results.

While the present studies did not use randomized samples, we suggest using a true experimental design for further studying GWPR instruction, because this can yield stronger evidence. An example of a true experimental design can be found in Kieft, Rijlaarsdam & Van den Berg (2006). The researchers assign students randomly to two experimental conditions, such that each class comprises students from both conditions. The material is self-instructing, and the lesson structure is identical for both groups to ensure that the differences between the tasks are not noticed by teacher candidates. For GWPR instruction, this design can be applied by administering writing tasks with different topics for learning for each of the two experimental conditions. Effects can be demonstrated by comparing topic knowledge and insight on the different topics that the teacher candidates were writing about.

The present think aloud studies were performed with very small samples. Although the analysis of protocols is very costly and time consuming, it is recommendable that future studies are carried out with more sizable samples. For substantiating the hypotheses provided in our study about the process of writing-to-learn, testing in larger samples is needed. It would provide more certainty about the issue whether the different types of reflection discriminated are indeed components of the process of writing-to-learn.

The two present studies provided evidence that certain elements of the (final) writing tasks have consequences for the process of writing-to-learn. While the final mathematics writing task explicitly defined an audience to write for, this element was missing in the final biology writing task. In addition, this task allowed teacher candidates to copy formulations from an existing text, which probably prevented them from critically reexamining their formulations. Therefore, in future studies final writing tasks should at least comply with the structure of writing tasks that are part of the GWPR instruction, including both a lay audience to keep in mind and the production of text that can be regarded as the teacher candidates' genuinely own text.

3.10 Pedagogical implications

Although it appears that writing-to-learn can be applied in many disciplines, it is not much used in education yet (Klein & Boscolo, 2016). Teachers are hesitant to use writing as a learning tool in class, as appeared from the interviews held with the teachers in the present studies. They felt insecure in designing writing-to-learn tasks and in selecting (or rather creating) good model texts as examples of the realization of certain conceptual relations. Additionally, they considered supporting student writing not as their job. It may seem self-evident that language teachers readily take on the task to support their colleagues on using writing-to-learn. However, this cannot be expected from them that easily. After all, their profession is not teaching writing-to-learn but learning to write, meaning that they instruct teacher candidates in how to structure their texts, connect sentences, use correct grammar and spelling. Most language teachers have no experience in composing model texts from a genre for writing-to-learn tasks and how to use these in class. Therefore, we suggest that language and subject teachers cooperate gaining experience in developing good writing-to-learn tasks. This way, all teachers can contribute their own expertise (on discipline specific knowledge or on writing tasks and instructions) and determine

which genre suits their learning goals. At the same time, they can use the elements of Genre Writing instruction: a preparatory activity for explaining the genre at stake, highlighting conceptual relations as the focus of instruction, followed by instruction on planning, reviewing and revising activities.

The present studies were of a small scale, but the results provide sufficient reason to continue research on GWPR. The studies provide new perspectives on writing-to-learn: the use of GWPR in largely differing disciplines leading to positive effects; the importance of cooperation between researchers and teacher educators, and attention to indicators of teacher candidates' writing-to-learn process made visible by think aloud studies.

Chapter 4

Effects of instruction in writing-to-learn on low-achieving adolescents in biology and mathematics classes

Chapter 4

Effects of instruction in writing-to-learn on low-achieving adolescents in biology and mathematics classes[3]

Abstract

This study investigates the effects of instruction in genre writing with planning and revising activities (GWPR) on learning. This type of instruction appeared to be successful in promoting learning in several types of education. However, there are few studies on the effects on low achievers. Therefore, two studies were conducted with low-achieving students, each comprising a quasi-experimental study and a small scale think aloud study, both of which were embedded in regular education for low-achieving adolescents. The first study took place in biology classes (grade 7, three lessons); the second study was in mathematics classes (grade 10, six lessons). The researchers co-created writing-to-learn tasks with the teachers. The results showed positive effects on learning in mathematics classes as compared with the control group, but not in biology classes. The think aloud study in the experimental mathematics class condition provided evidence of the learning by writing process. In the experimental biology class condition, such evidence was barely present. The results suggest that the experimental intervention in biology classes was too short for the students to grasp the essentials of learning by writing. This paper also discusses suggestions for further research and pedagogical implications.

[3] This chapter is a slightly modified version of the published article:
Van Dijk, A., Van Gelderen, A., & Kuiken, F., (2022). Effects of instruction in writing-to-learn on low achieving adolescents in biology and mathematics classes. *Education Sciences*. https://doi.org/10.3390/edusci120595

4.1 Introduction

The present study investigates the effects of instruction in writing-to-learn in two disciplines. Writing-to-learn (or: learning by writing) means that students carry out writing tasks intended to stimulate reflection on their knowledge about a disciplinary topic, which may lead to new insight and topic knowledge (Bereiter & Scardamalia, 1987; Galbraith, 2009; Van Dijk et al., 2022a). Topic knowledge refers to basic factual knowledge, such as names and dates. By insight, we mean the ability to relate new concepts to students' prior knowledge.

Over the years, writing-to-learn in educational contexts has been studied frequently. Researchers observed that teaching often consisted of mere transmission of knowledge and rarely stimulated students to arrive at new insights (e.g., Britton, 1982; Langer & Applebee, 1987; Mason, 1999). Therefore, writing-to-learn was suggested as a promising addition to traditional teaching.At first, researchers held varying views on how to stimulate the process of writing-to-learn and research showed mixed results (Klein, 1999). For instance, the "Writing Across the Curriculum" (WAC) movement suggested that using elements of good writing, such as attention to text organization and conciseness, leads to students' insight in subject matter (Klein, Boscolo, Gelati & Kirkpatrick, 2014). Bazerman (2005), however, observed that each discipline entails a specific way of reasoning, which means that a disciplinary approach of writing-to-learn is required

Recent research has shown that writing can lead to positive effects on insight and topic knowledge (Klein & Boscolo, 2016; Graham et al., 2020; Miller et al., 2018; Hand et al., 2021; Van Dijk, et al., 2022). Studies are directed at students in all levels of education (from primary to higher education). However, few studies have been specifically aimed at writing-to-learn interventions for low-achieving adolescents. Although many studies may have included low achievers, results have rarely been reported specifically for that group. Therefore, it is not yet known whether low achievers can be supported by writing in the same way as other students. As Rivard (2004) states, it may be that low achievers need a different type of writing-to-learn instruction compared to high achievers. Therefore, we set out to bolster the scant research specifically directed at low-achieving students in secondary education.

We investigate whether instruction in writing-to-learn can be beneficial for low-achieving students. On the one hand, there may be reasons why it is not. For instance, low-achieving students often have difficulty with the "basics" of the writing process (such as formulating correct sentences and spelling), and therefore have too little cognitive room to focus on text contents and reflect on disciplinary topics. In addition, low achievers may be less motivated to write in comparison to students with higher achievement levels, which may profoundly inhibit the conditions for writing-to-learn.

On the other hand, there is reason to believe that instruction in writing-to-learn is effective for low-achieving students as well, because writing facilitates reflection on one's own thoughts by externalizing these thoughts in written text (Nückles, Hübner & Renkl, 2009). Low-achieving adolescents may profit from this externalizing of their own ideas in ways similar to those of other students. We investigate whether instruction in specific writing-to-learn tasks in two disciplines

(biology and mathematics) improves low-achieving adolescents' topic knowledge and insight and whether we can identify indicators of the process of learning by writing in their task approach.

4.2 Research into instruction in writing-to-learn

In recent research, explicit instruction is regarded as conditional for stimulating writing-to-learn. Bangert-Drowns et al. (2004) were the first to highlight this point-of-view. They conducted a meta-analysis based on 48 studies in primary, secondary and higher education, concluding cautiously that instruction in cognitive and metacognitive strategies may lead to learning by writing. Cognitive writing strategies are understood as organizing strategies, such as goal setting, selecting, and structuring contents. Metacognitive writing strategies are understood as strategies for monitoring task performance and evaluating texts (reviewing and revising). Bangert-Drowns et al. (2004) assumed that such instruction stimulates reflection on writing products, which may lead to new insight and topic knowledge .

More recent reviews of research into writing-to-learn confirmed that instruction in cognitive and metacognitive strategies may stimulate learning by writing (Graham et al., 2020; Miller et. al., 2018; Hand et al., 2021, Van Dijk et al., 2022a). In their meta-analysis of 56 (quasi) experiments, Graham et al. (2020) found positive effects of explicit instruction in cognitive and metacognitive strategies for writing-to-learn in 82% of (quasi) experiments in science, social studies, and mathematics in grades 1–12. However, the researchers could not determine which elements of instruction were most effective in promoting learning.

A systematic review of 43 studies by Miller et al. (2018) took place in regular courses in grades 6-12 in science, social studies, and mathematics.The researchers concluded that three types of instruction resulted in positive effects on learning. The first type was explicit instruction of cognitive strategies. The second was inquiry-based instruction stimulating students to discover how to use cognitive and metacognitive strategies without guidance by the teacher. The third was instruction of metacognitive strategies for students' reflection on their learning. Overall, in 46.5% of reviewed studies, instruction consisting of one of these types had positive effects on learning.

Hand et al. (2021) reviewed 81 theses (41 quantitative and 40 qualitative) on the use of the Science Writing Heuristic (SWH) in secondary education. The SWH is a tool for instructing cognitive strategies for planning (organizing and generating) a science text. The researchers found that its use in the quantitative studies resulted in the growth of insight regardless of grade or cultural background. Additionally, a qualitative study on elements maximizing effects on learning showed that a determining factor is the amount of time available. This appeared to be true for students when engaged in using the SWH, as well as for teachers gaining experience with it.

Van Dijk et al. (2022a) reviewed 43 studies in primary (from grade 5), secondary, and higher education in all disciplines. The authors distinguished four types of instruction, three of which were based upon hypotheses about the process of writing-to-learn proposed by Klein (1999), whereas the fourth type emerged from a

considerable amount of reviewed studies. The first type is Forward Search, which consists of metacognitive strategies for revising the contents of a draft. The second is Genre Writing, which provides students with a model of the assigned genre, sometimes supplemented by cognitive and/or metacognitive strategies for planning and revision. The third is Backward Search, which means including instruction of cognitive strategies for planning and metacognitive strategies for revising a draft. The fourth is Planning Only, which is instructing cognitive strategies for planning. Results showed positive effects on insight and topic knowledge for all four types. In total, positive effects were found in 66% of the experimental studies.

Apart from the cognitive and metacognitive strategies involved in writing-to-learn that emerged from the above research syntheses, there are also some important conditions mentioned in the literature. Newell (1984) considered genre knowledge conditional for writing-to-learn. If writers are not familiar with conceptual relations characteristic of an academic genre (for instance, causality in expositions), writing a text in that genre becomes extremely difficult. In addition, Klein and Kirkpatrick (2010), and Klein and Samuels (2010) concluded that genre knowledge is conditional for writing-to-learn. In the review by Van Dijk et al. (2022a), Genre Writing appeared as one of the main types of instruction. Genre Writing refers to teachers' explanation of characteristic ways conceptual relations are accomplished in the given genre. Another important condition for writing-to-learn concerns the audience writers are targeting. Prain (2006) suggested that instructing students to write for a lay audience requires them to reflect more deeply on how to reformulate the academic language from their textbooks into everyday language. This reflection and reformulation may lead to new insight and topic knowledge.

The reviews by Graham et al. (2020) and Miller et al. (2018) reported small numbers of studies investigating writing-to-learn by students of various abilities. Graham et al. (2020) noted that in two of 56 experiments, the participants were low-achieving adolescents, with both showing positive effects. Miller et al. (2018) mentioned one study leading to positive effects. Hand et al. (2021) mentioned two studies not showing significant differences between low achievers and high achievers on a post-test when compared to each other. In the review by Van Dijk et al. (2022a), no studies were encountered focusing exclusively on low-achieving students.

4.2.1 Low- and high-achieving students in research on writing-to-learn

Indications of differences in the usefulness of writing-to-learn activities for higher- and lower-achieving students can be found in only a few studies. Rivard (2004) investigated which specific language activities were beneficial for learning for 154 low- and high-achieving adolescents (eighth grade) in science class. The study compared three experimental groups to a control group, each group comprising high- and low-achieving adolescents. One group was instructed to carry out a task by means of peer discussion, another by means of writing preceded by peer discussion, and the third by writing only. The three groups were compared to a control group that performed restricted writing tasks (such as matching exercises, true-or-false questions, or fill-in-the-blanks). Low-achieving students appeared to acquire more

topic knowledge after performing writing activities preceded by peer discussion about content selection and organization of their text. High achievers benefited the most from writing activities only.

Akkus, Gunel, and Hand (2007) compared 322 experimental with 270 control students in science education, each condition comprising of (seventh to eleventh grade) high- and low-achieving students. Experimental students were instructed to perform inquiry activities by means of a checklist and to discuss these with peers before writing, just as Rivard (2004) did. Control students listened to the teacher's explanation of subject matter before writing. The researchers found that experimental low-achieving adolescents showed larger effects on insight and topic knowledge than control students. The explanation by Akkus et al. (2007) is that peer discussion taking place in everyday language rather than academic language, which is usual in class, may support low achievers in understanding subject matter. High-achieving students gained equally high scores in both conditions; thus, the type of instruction did not matter.

Faber, Morris, and Lieberman (2000) noticed that low-achieving students appeared to struggle when instructed to write notes for understanding their textbook, whereas high-achieving students did not. Therefore, they investigated whether a preparatory training on note taking for reading textbook sections led to positive effects on low-achieving students' learning. The study compared 112 experimental ninth grade students receiving note-taking training before studying their textbooks, with 90 control students receiving the business-as-usual program of the subject World Cultures. Both groups of students consisted of high- as well as low-achieving students. The activities preparing experimental students on note taking comprised the training of reading and note-taking skills. Low- as well as high-achieving experimental students showed more insight and topic knowledge into subject matter than control students.

The three studies above support the idea that low-achieving students may benefit from instruction using writing-to-learn tasks. Peer discussion for planning (selecting and organizing) or a preparatory activity on note taking appeared to lead to positive effects on low-achieving students' acquisition of topic knowledge Rivard (2004), or on topic knowledge as well as insight in different disciplines. Therefore, it makes sense to further investigate the effects of instruction in writing-to-learn for low-achieving adolescents.

4.3 Supporting low-achieving students for learning

In our study, we combined the elements from the literature that appeared promising for stimulating writing-to-learn for low-achieving adolescents. We called the result Genre Writing instruction combined with Planning and Revising (GWPR). The elements are the following: genre knowledge, cognitive and metacognitive strategies, peer discussion, and writing for a lay audience. GWPR instruction provides low-achieving students with support at three phases of the writing process: before planning, during planning, and during reviewing.

First, GWPR provides genre knowledge in a preparatory activity. The genre is explained by means of a model text written in the given genre. Genre knowledge is defined as knowledge of the genre's rhetorical goal and prevalent conceptual relations between text elements needed to arrive at this goal (Halliday & Martin, 1993). For instance, the rhetorical goal of the genre explanation is clarifying, and a conceptual relation to arrive at this goal is 'condition' (e.g., in mathematics, 'If you know how many data you have, you can determine the size of your table'). Students need genre knowledge for understanding their textbooks and for relating concepts in their own texts. Therefore, the model text should illustrate how conceptual relations can be linguistically realized, for instance by means of 'because' or 'therefore' for relating a statement to an argument, or by means of 'first'...', then ... ', 'finally...' for relating steps in an instruction. By reflecting on composing a conceptual relation, for instance an argumentative relation, students may acquire new insights into the meaning of the conceptual relation (Newell, 1984; Klein & Kirkpatrick, 2010; Klein & Samuels, 2010). The model text supports students in writing a text in a given genre. Additionally, the model text is directed at a lay audience, demonstrating how to formulate the conceptual relations in a non-academic fashion for readers with little prior knowledge of the subject.

Second, GWPR contains the cognitive strategies selecting and organizing needed for planning how to write the text. For selecting contents, students use their memory and their textbooks. For organizing contents, they decide how to relate selected contents by means of the previously demonstrated conceptual relations. Students discuss their planning in pairs, implying that they evaluate the appropriateness of each other's selection of contents for the intended lay audience as well as the comprehensibility of the conceptual relations. Peer discussion on planning supports them in reflecting on how to relate selected concepts and to organize these in a text. Students' reflections may result in new insights (Bereiter & Scardamalia, 1987). Additionally, students' discussion about the accessibility of their texts to their audience may lead to new insights (Prain, 2006) . The model text supports students in directing their texts at a lay audience.

Third, after students have written their drafts individually, they review their peers' drafts by reflecting on the description of conceptual relations and comprehensibility to the lay audience. Next, peers reflect on each other's feedback and ask for clarification, if needed. Then, students revise their drafts individually using the received feedback requiring them to reflect on their original ideas, which may lead to new insights (Bereiter & Scardamalia, 1987).

Finally, writing to a lay audience is inherent in GWPR in the planning and revising phases. It entails that writers cannot use the academic language from their textbooks for an audience that has little prior knowledge about the topic. Instead, they must use everyday language to make their text comprehensible.

4.4 The process of writing-to-learn

Only a few studies examined the process of writing-to-learn by using think aloud studies or keystroke logging for analyzing students' thinking. One study took place in elementary education (Bereiter & Scardamalia, 1987), and four were conducted in

higher education (Galbraith, 2009; Baaijen & Galbraith, 2018; Klein, 2004; Van Dijk et al., 2023). Low achievers were not involved in these studies. In our study, we decided to add to the existing evidence using empirical data for analyzing low-achieving adolescents' writing-to-learn processes.

Two theories about the nature of the writing-to-learn process were proposed, one by Bereiter and Scardamalia (1987) and one by Galbraith (2009). Bereiter and Scardamalia (1987) assume that the writing process of experienced writers is a recursive process between contents and rhetorical goals for planning, reviewing, and revising [2]. Writers seek to reconcile contents and rhetorical goals, and therefore adjust their text several times on rhetorical and content aspects, resulting in knowledge transformation. By performing such activities, writers acquire new insights. According to this theory, learning by writing takes place in a cyclic process entailing reflection on content and rhetorical goals interactively.

Galbraith (2009) proposed the dual process theory that distinguishes between a knowledge retrieval process and a knowledge constituting process. Writers use their knowledge retrieval system for retrieving content while considering rhetorical goals. While writing, they use their knowledge constituting system to make connections between concepts, some of which may be connections the writer was not aware of previously (implicit knowledge). These new connections lead to new insights. The constitution of connections is a cyclic process in which writers alternately revise their text and refer to their knowledge constituting system. The two theories have in common that learning by writing takes place in a cyclic process entailing reflection on and adaptation of content and rhetorical goals.

Bereiter and Scardamalia (1987) tested their theory in an empirical study with sixth grade students. They considered reflection as indicating knowledge transformation. In a think aloud study comparing six experimental with six control students, they found evidence that experimental students progressed more in transforming their knowledge during the planning of their texts in comparison to the control group.

Recently, Baaijen & Galbraith (2018) used keystroke logging for measuring 78 university students' writing processes as described in the dual process theory. Students' revision of text, which entailed reflection, appeared to be related to increased insight.

In a think aloud study with 56 university students aimed at identifying writing-to-learn processes by conducting an exploratory think aloud study, Klein (2004) confirmed the outcomes by Bereiter & Scardamalia (1987) that reflection on planning activities, such as goal setting, organizing, and generating, appeared to promote insight. Additionally, Klein (2004) reported that reflection on reviewing and revising also resulted in insight. Van Dijk et al. (2023) found that experimental biology and mathematics teacher students receiving GWPR instruction showed more reflection on planning as well as on reviewing and revising than a control group. Thus, Klein (2004) and Van Dijk et al. (2023) appear to find support for both theories. Klein (2004) concluded that positive effects of writing on learning might result from both knowledge transformation and knowledge constitution.

We investigate the writing-to-learn process of low-achieving students in secondary education to find out whether reflection on planning, reviewing, and

revising are involved, just as they are for other student populations. In doing so, we intend to show how GWPR instruction contributes to low achievers' learning.

4.5 The present studies in biology and mathematics education

The present studies investigate whether GWPR instruction leads to increased insight and topic knowledge of low-achieving adolescents. Additionally, the studies aim to identify indicators of the writing-to-learn process. The studies were carried out in the context of Dutch prevocational education. This educational track is aimed at students with the 30% lowest scores on an academic aptitude test (language and mathematics) administered before the start of secondary education (in grade 6). Therefore, these students are characterized as low achieving.

Pre-vocational education appeared to be a relevant context because writing-to-learn can be an appropriate way of learning for students in this type of education, providing them with opportunities for more active understanding and participation in academic discourse by writing. In addition, the teachers involved were fourth (last) year apprentice teachers, who had participated in a previous writing-to-learn experiment as students. Their class had served as experimental condition in a study on effects of GWPR instruction in teacher education (Van Dijk et al., 2023). Therefore, we could take advantage of the apprentices' experience and knowledge about writing-to-learn in designing the interventions and adapting them to the target group of students.

The studies were situated in two subject areas: biology and mathematics. This allows us to compare results from two widely different subject areas, in regard to the role that writing plays in regular educational practice. In biology, writing is a regularly used activity (for instance, for executing tasks from the textbook), whereas in mathematics, writing rarely takes place (Veel, 1999).

We report two studies, each comprising a quasi-experiment, one in biology (Study 1) and one in mathematics class (Study 2). GWPR was applied in both studies to be able to compare effects between both disciplines. Following recommendations by Hand et al. (2021), the interventions were embedded in regular lesson series that were part of the curricula for biology and mathematics. The studies were carried out in co-creation with the teachers for organizing, composing, and embedding the intervention in the regular courses.

Additionally, think aloud studies were included in Study 1 and Study 2. The combination of a quasi-experiment and a think aloud study offers the opportunity to investigate whether experimental students not only acquire more knowledge and insight but also shows specific indicators of learning in their writing process. The process of learning while writing was analyzed by comparing samples from the control and the experimental group.

We formulated two research questions:

1. Does GWPR instruction lead to more insight and topic knowledge of low-achieving adolescents in the context of biology and mathematics when compared to business-as-usual lessons?

2. Does GWPR instruction in the above contexts lead to observable differences in processes of writing-to-learn between experimental and control students?

4.6 Study 1 Writing-to-learn for low achieving students in biology class

4.6.1 Materials and Methods

4.6.1.1 Participants

The study took place in May and June 2014. Four seventh grade classes of two pre-vocational schools in two cities participated. Classes within a school were randomly assigned to an experimental and a control condition. From the start, 102 students were present in these classes (50 control and 52 experimental). A number of 96 students (46 control and 50 experimental students) completed a (required) prior knowledge and vocabulary test at the start of the experiment. Six students were absent. Two other students were absent during the administration of the post-test, resulting in a sample of 94 students. Because of technical problems in one school during the data collection, students' answers on the post-test were incomplete. Therefore, we were not able to analyze the data of all 94 participants. This resulted in a final sample of 75 participants (32 control and 43 experimental). Table 4.1 presents the age, gender, and mother tongue of the final sample.

For answering the second research question about the process of writing-to-learn, 10 students were randomly selected from the sample of 75 students, five students belonging to the control group and five to the experimental condition. Table 4.2 shows characteristics of students.

Table 4.1 Characteristics of the participants

	Experimental group (N=43)	**Control group (N=32)**
Age	M: 12.8 (SD: 0.74)	M: 12.8 (SD: 0.61)
Gender	Female: 21	Female: 15
Mother tongue	Dutch: 29	Dutch: 22

Table 4.2 Characteristics of the selected participants of the think aloud study

	Experimental group (N=5)	**Control group (N=5)**
Age	M: 12.4 (SD: 0.55)	M: 12.6 (SD: 0.55)
Gender	Female: 4	Female: 3
Mother tongue	Dutch: 3	Dutch: 2

Two biology teacher students completing their final internship in two schools were involved as teachers in the study. In one school, the teacher student instructed the experimental condition while an experienced colleague instructed the control group. In the other school, the teacher student taught both the experimental and control groups. The teacher students were fourth (final) year students at the teacher education institute of a University of Applied Sciences. In the third year of their study, they had

participated in an experiment (Van Dijk et al., 2023) directed at the effects of GWPR instruction in biology teacher education. Therefore, they were familiar with our approach of instruction in writing-to-learn and motivated to adapt this approach to the teaching context of prevocational education.

4.6.2 Design

We used a quasi-experimental, post-test only design, with a business-as-usual control group and an experimental group. The dependent variables were insight and topic knowledge into biology subject matter taught. Prior knowledge of biology and vocabulary knowledge were used as covariates. The control group received regular lessons (without writing tasks) while the experimental group received GWPR lessons comprising writing tasks aimed at writing-to-learn.

For analyzing the process of writing-to-learn, a think aloud multiple case study was carried out with five experimental and five control students, randomly selected from the final sample. A post-test writing task was used for the think aloud procedure (see Instruments). Utterances were coded and systematically analyzed to investigate differences between the writing processes of experimental and control students.

4.6.3 Treatment

The experimental intervention took place in three 45-min lessons. The lessons were part of a series of eight lessons aimed at insight and topic knowledge into the musculoskeletal system. Preceding the experiments, the first author observed regular lessons in both schools to determine their structure in preparation for composing the intervention (more specifically, how to insert the writing-to-learn tasks). The lesson structure appeared to be identical in both schools.

The first author cooperated with the two teacher students in designing three writing-to-learn tasks, accompanying model texts and a teacher's manual. First, they determined which lesson series from the textbook (both teachers used the same) was most suitable for embedding the writing tasks. Then, they discussed which parts of the lessons to replace with writing tasks. Finally, the first author proposed topics for the three writing tasks and the model texts, which were evaluated by the teachers on how these were related to the main aims and contents of the lesson series.

It was decided to focus on the conceptual relations 'purpose and means', 'comparing', and 'cause and effect' because these relations appeared to be frequently used in the regular lessons. These conceptual relations match the genre 'explanation' (Rose, 2008). Therefore, the writing tasks required students to write explanatory texts, each focusing on a different conceptual relation. GWPR instruction entails that each writing task is preceded by a teacher's explanation of a model text. Three model texts were derived from a textbook directed at the target audience of the writing tasks (grade 6) and rewritten to fit our needs. The topics of the model texts were related to that of the writing task, such that students could use the model as an example for how to realize the conceptual relations. However, the topics of the model texts were not identical with the writing tasks to avoid copying. For instance, the topic of a model text was 'using muscles for kicking a ball (soccer)' and the topic of the writing task was 'using muscles for throwing a ball (basketball)'. Each model text contained one

conceptual relation (as in the writing task) that was expressed in various ways (e.g., the relation 'purpose and means' was expressed as: 'use… to … '; 'for that … you use... '). The purpose was to familiarize students with various linguistic expressions of a conceptual relation.

In each writing task, the experimental participants were instructed to use the conceptual relation present in the model text when writing their explanation. Furthermore, students were instructed to direct their text at a younger audience of sixth grade students. By carrying out such tasks, students needed to reflect on how to transform their own thinking to adjust to their audience (the rhetorical plane of writing). We assume that this may lead to students' awareness of the need for a clear (and simpler) explanation of relations. GWPR assumes that this awareness results into more explicit topic knowledge and insight into subject matter (Prain, 2006).

The GWPR writing tasks consisted of the following parts. The first concerned pairwise planning activities by brainstorming on possible contents. Students had to take into account the assigned conceptual relation and the intended audience. The second part involved the writing of a draft individually. The third required students to give feedback on each other's draft referring to the conceptual relation and the intended audience. Finally, students had to revise their drafts using the received feedback.

The teachers received a manual for how to instruct students to do their homework and how to present the model texts and the writing tasks to the students. It contained suggestions for explaining the conceptual relation to the students, with the model text projected on a smart board, highlighting various linguistic expressions of the conceptual relation. In addition, students were asked to look for other examples of the conceptual relation in the model text and discuss these.

4.6.4 Instruments

4.6.4.1 Prior knowledge tests

In consultation with the two teacher students, the first author composed the tests for determining students' prior topic knowledge and insight. The tests were based upon subject matter taught in the first part of the seventh-grade academic year as well as in grade 6. The grade 6 curriculum contains several different themes related to biology, varying from nature to sports. Eight items were derived from the biology textbook for grade 7 (Smits, 2008) and five were derived from a national test for investigating sixth grade students' knowledge level in biology (Thijssen et al., 2011). Four items around the theme 'sports' were composed in collaboration by the first author and the teacher students. The prior topic knowledge test consisted of 10 multiple choice items; the prior in-sight test consisted of seven open-ended questions. Two items were deleted from the prior topic knowledge test because one showed no variability in students' answers, and the second item was invalid, thus eight items were left. We did not compute interrater reliability for the open-ended questions because these were unambiguous. (For instance, 'What kind of clothes are best to practice sports in?').

The homogeneity of the tests measuring prior topic knowledge and insight is low (Cronbach's alpha is 0.36 and 0.34 respectively), which was expected because of the large variation in biological themes tested. Cronbach's alpha provides an underestimation of test reliability (Boyle, 1991; Sijtsma, 2009; Taber, 2018).

Therefore, the tests might still be sufficiently reliable to explain variance in our posttest measures. Consequently, we decided to include both tests as covariates in our analysis.

4.6.4.2 Vocabulary

A vocabulary test of 30 items derived from the Dutch version of the Peabody Picture Vocabulary Test (Dunn & Dunn, 2005) was composed. This Dutch version is based on frequencies per one million words. The words are ranked in 17 sets, each aimed at a specific age group. For the selection of words, we used three sets (nr. 6, 7, 8) aimed at ages below 15. We selected words of which the expected proficiency was $p = 0.50$ and higher. Nine items did not show variability and were therefore deleted. Cronbach's alpha over 21 items was 0.68, which is regarded as acceptable.

4.6.4.3 Topic knowledge and insight in the post-test

The post-test consisted of 14 multiple choice items (topic knowledge) and 10 open-ended questions (insight). The items and questions were derived from a test that was part of the biology textbook used in grade 7. Cronbach's alpha of the topic knowledge test was 0.20. Because of this low homogeneity, this test was not included in the analysis. Unlike the tests for prior knowledge, the topic knowledge post-test consisted of items from the narrow domain of lesson contents about the musculoskeletal system. Therefore, a much higher homogeneity of the topic knowledge test was expected here.

Cronbach's alpha for the post-test for insight was 0.70, which is acceptable given the narrow domain from which the items were selected. Therefore, the test was included in our analysis. The open-ended questions were formulated in such way that only one answer was possible (for instance: which type of joint is in a knee?). Therefore, there was no need to calculate inter-rater reliability.

4.6.4.4 Post-test writing task

Finally, a post-test writing task was administered. Students were asked to write an explanation about the correct body posture for lifting a heavy weight. The explanation had to be directed at grade 6 students entering high school the following academic year. We used this writing task for examining the process of writing-to-learn in the think aloud study.

4.6.5 Procedure

Biology class was scheduled twice a week. The study involved six lessons of 45 min each. The first lesson was used for administration of the tests for prior insight, topic knowledge, and vocabulary; the fifth lesson was used for the post-test (topic knowledge and insight); and the sixth lesson for the post-test writing task. Lessons 2–4 were dedicated to the three writing tasks for the experimental condition and the business-as-usual lessons in the control condition.

Table 4.3 presents the lesson structure for these three lessons in the control and experimental condition. The lesson structure in the control condition consisted of explanation of new theory and the completion of textbook assignments. For homework, students were required to finish assignments from the textbook that had not been completed. The three experimental lessons each comprised one writing-to-

learn task consisting of an explanation of the model text, followed by planning and writing a first draft, and giving and receiving feedback for writing a final draft. In each lesson, experimental students received a writing task about theory they had studied as homework prior to the lesson.For securing treatment fidelity, the first author observed whether the teacher students in both schools carried out the instructions conforming the teacher manual. No deviations were encountered in that respect. However, in both schools, the observations also showed that students in the experimental condition at first worked seriously but became less concentrated when giving feedback to each other and revising their drafts. Therefore, the first author discussed this with the teacher students aiming to improve students' engagement. For instance, teacher students explained that students' feedback contained suggestions for replacing wordings that were incomprehensible for the intended audience. This served as a boost for the students' motivation to put effort in the review and revision of their texts.

For keeping time on task equal for both conditions, the elements 'explanation of new theory' and 'performing assignments from the textbook' were replaced in the experimental group by respectively explanation of the model text and planning, writing, feedback, and writing final draft (see Table 4.3). The homework 'completing unfinished assignments' was replaced by 'studying new theory', which was assigned in lesson 1. The teacher explained the experimental students how to study new theory individually by means of a roadmap on reading strategies because they were not used to performing such task.

Table 4.3 Lesson structure in control and experimental conditions

Control group **Business as usual lessons 2, 3, 4**	**Experimental group** **Writing-to-learn tasks lessons 2, 3, 4**
• Explanation of new theory • Performing assignments from the textbook	• Explanation of a model text • Planning and writing a first draft • Feedback and writing a final draft
Homework:completing unfinished assignments	Homework: studying new theory

For investigating the process of writing-to-learn, a random sample of 10 students (five experimental and five control) performed the writing task thinking aloud. The sessions took place in lesson 6. These students performed their writing tasks individually in an empty classroom in the presence of the first author. She said that she was interested in how they addressed the writing task; for this reason, she asked them to think aloud while writing. She provided an instruction comprising a video clip of a student thinking aloud while writing to demonstrate what was expected from the students. When students kept silent for 10 s, the first author encouraged them to keep thinking aloud, and used prompts such as: please, say aloud what you are thinking. Students wrote their texts by hand. The duration of the sessions was 15–20 min. The sessions were video recorded.

4.6.5.1 Coding Students' Transcribed Utterances

Students' verbalizations were transcribed and represented in protocols as separate utterances in cases of verbal behavior and separate instances in cases of non-verbal behavior (for instance, frowning). An utterance was defined as a phrase containing one complete element of information (Pander Maat, 1994).

The codes were based on Hayes and Flower's (1980) writing model. The following writing activities were distinguished: (1) planning, including four subcategories (orienting, generating, selecting and organizing); (2) formulating; (3) monitoring; and (4) evaluating. In total, 24 codes to describe students' writing and thinking processes were used. For instance, the utterance 'yes, this should be it' was coded as 'selecting content'. In addition, interruptions and utterances not focusing on the writing task were coded as 'other activities'. Finally, the resulting coding scheme comprised 29 verbal and non-verbal activities divided over five main categories (see Appendix C). By means of this coding scheme, a researcher and a research-assistant coded the utterances, one code per utterance or per instance. For determining inter-rater reliability, two protocols (one for each condition) were coded by two independent raters. There was agreement for 84% of all utterances/instances, which is acceptable for our purposes. Differences in coding were resolved after discussion.

Reflective activities such as reviewing, revising, goal setting, organizing, and generating contents can be regarded as indicators of writing-to-learn processes, according to previous studies (Bereiter & Scardamalia, 1987; Baaijen & Galbraith, 2018; Klein 2004). The following specific codes from our list are regarded as indicators for these reflective activities: (1) using knowledge about audience; (2) thinking about content selection; (3) thinking about formulating; (4) revising while formulating; (5) revising after finishing an utterance; (6) rereading own text; and (7) rethinking task approach.

4.6.5.2 Data Analysis

Three analyses of variance (ANOVA's) were used for comparing prior topic knowledge, prior insight, and vocabulary in the two conditions. By means of analyses of covariance (ANCOVA), students' post-test scores on insight in the two conditions were compared. Prior insight, prior topic knowledge, and vocabulary were used as covariates. In all statistical tests, alpha level was 0.05.

Utterances indicative for the process of writing-to-learn and other utterances were systematically analyzed to investigate differences between the experimental and control students that were randomly selected from the sample. To determine whether indicators for writing-to-learn occurred more often in the experimental group than in the control group, the means of each code per condition were computed. Subsequently, the ratio of the mean frequency of each code to the total number of utterances in each condition was computed. Finally, effect-sizes (Cohen's d) were computed for estimating the magnitude of differences between conditions on these proportioned mean frequencies (.20: small, .50: medium, .80: large).

4.6.6 Results

Table 4.4 presents the means and standard deviations for the four variables involved, i.e., prior insight, prior topic knowledge, vocabulary, and post-test insight.

Table 4.4 Means and standard deviations for prior insight, prior topic knowledge, vocabulary knowledge, and post-test insight

	Experimental group (N=43)	Control group (N=32)
Variables	**Mean (SD)**	**Mean (SD)**
[1)] Prior insight	5.65 (1.71)	5.50 (1.92)
[2)] Prior topic knowledge	5.60 (1.31)	5.56 (1.64)
[3)] Vocabulary	19.65 (1.15)	19.41 (2.58)
[4)] Post-test insight	6.16 (2.46)	6.84 (2.43)

[1)] theoretical maximum score: 7 [3)] theoretical maximum score: 21
[2)] theoretical maximum score: 8 [4)] theoretical maximum score: 10

4.6.6.1 Prior insight, prior topic knowledge, and vocabulary

ANOVA showed that experimental students' and control students' scores on prior insight did not differ significantly: $F(1,73) = .13$, $p = .72$, partial $\eta^2 = .00$. Scores of the two groups on prior topic knowledge did not differ significantly either: $F(1,73) = .02$, $p = .90$, partial $\eta^2 = .00$. Finally, the scores of the two groups on the vocabulary test did not differ significantly: $F(1,73) = .31$, $p = .58$, partial $\eta^2 = .00$.

4.6.6.2 Differences in post-test insight

Two ANCOVAs were performed to compare experimental students' and control students' scores on post-test insight. In the first ANCOVA, prior topic knowledge, prior insight, and vocabulary served as covariates. Prior topic knowledge and prior insight predicted the scores on post-test insight significantly: $F(1,73) = 5.09$, $p = .03$, partial $\eta^2 = .07$, and $F(1,73) = 5.85$, $p = .02$, partial $\eta^2 = .08$. The remaining covariate (vocabulary) did not predict the scores significantly: $F(1,73) = 1.25$, $p = 0.27$, partial $\eta^2 = .02$. Therefore, in the second ANCOVA, the latter covariate was omitted. The analysis showed again that prior topic knowledge predicted post-test insight significantly: $F(1,73) = 5.34$, $p = .02$, partial $\eta^2 = .07$; the same was true for prior insight: $F(1,73) = 7.63$, $p = .01$, partial $\eta^2 = .10$. However, experimental students did not differ significantly from control students on the post-test insight: $F(1,73) = 2.51$, $p = .12$, partial $\eta^2 = .04$.

4.6.6.3 The Process of writing-to-learn

Table 4.5 shows the proportioned means and standard deviations of the seven codes representing reflective activities that are considered indicators of the writing-to-learn process (see Appendix A for an overview of all 29 codes). The second, third, fourth, and fifth columns show the proportioned mean frequencies and standard deviations of codes for the experimental and control condition. The final column shows the effect size (Cohen's d).

Table 4.5 Proportioned means, standard deviations, and effect sizes of indicators of writing-to-learn in experimental (N=5) and control groups (N=5)

Codes	Experimental Mean (SD)	Control Mean (SD)	Effect size Cohen's d
PLANNING: generating			
Using knowledge about audience	.00 (.00)	.004 (.009)	.00
PLANNING: selecting			
Thinking about content selection	.09 (.07)	.02 (.03)	1.30
FORMULATING			
Thinking about formulating	.01 (.02)	.01 (.02)	.00
Revising while formulating	.005 (.01)	.02 (.03)	- .89
Revising after finishing an utterance	.01 (.01)	.01 (.01)	.00
MONITORING			
Rereading own text	.02 (.01)	.02 (.03)	.00
Rethinking task approach	.01 (.01)	.01 (.02)	.00

Table 4.5 shows that only one reflective activity, thinking about content selection, was used more frequently by experimental students than control students, d = 1.30 (large effect). The remaining activities did not show differences in favor of the experimental condition. However, control students performed one reflective activity, revising while formulating, more often than experimental students (large effect). Students in both conditions appeared to be rarely inclined to reflect on their writing on most of the indicators in Table 5.

4.7 Study 2 Writing-to-learn for low achieving students in mathematics class

4.7.1 Materials and Methods

4.7.1.1 Participants

The study took place in four tenth grade mathematics classes in two schools for pre-vocational education in two cities in November and December 2014. Classes within a school were randomly assigned to an experimental and a control condition. From the start, 74 students (33 experimental and 41 control) participated in the study. Seven students were absent when the post-test was administered. Therefore, the final number of students participating in the study was 67 (30 experimental and 37 control students). Table 4.6 presents students' ages, genders, and mother tongues.

For answering the second research question about the process of writing-to-learn, eight students were randomly selected from the sample of 67 students, four of whom belonged to the control group and four to the experimental condition. Table 4.7 shows the characteristics of these students.

Two mathematics teacher students were involved in the study, one in each school. In both schools, the teacher students instructed students in the experimental condition, while an experienced colleague instructed students in the control group. The teacher students were fourth year students at a teacher education institute of a university of applied sciences. In the previous year, they had participated in an experiment examining the effects of GWPR instruction in mathematics teacher education (Van Dijk et al., 2023) and had carried out writing-to-learn tasks themselves. Therefore, they were familiar with our approach of instruction in writing-to-learn and motivated to adapt this approach to the context of pre-vocational education.

Table 4.6 Characteristics of the participants

	Experimental group (N=30)	**Control group (N=37)**
Age	M: 15.8 (SD .57)	M: 15.8 (SD .64)
Gender	Female: 19	Female: 16
Mother tongue	Dutch: 27	Dutch: 35

Table 4.7 Characteristics of the selected participants in the think aloud studies

	Experimental group (N=4)	**Control group (N=4)**
Age	M: 15.3 (SD .50)	M: 15.5 (SD .58)
Gender	Female: 2	Female: 1
Mother tongue	Dutch: 4	Dutch: 4

4.7.2 Design

In a quasi-experimental, post-test only design, a control group and an experimental group were compared. The control group received business-as-usual lessons (without writing tasks) and the experimental group received GWPR lessons comprising writing tasks aimed at writing-to-learn. The dependent variables were insight and topic knowledge into the taught mathematics subject matter. Prior insight, prior topic knowledge in mathematics, and vocabulary knowledge were used as covariates.
For analyzing the process of writing-to-learn, a think aloud multiple case study was carried out with four experimental and four control students who were randomly selected from the total sample. A post-test writing task was used for the think aloud procedure (see Instruments). Utterances were coded and systematically analyzed to investigate differences between the writing processes of experimental and control students.

4.7.3 Treatment

The experiment took place in a lesson series aimed at insight and topic knowledge into mathematical relations. Preceding the experiments, the first author observed regular mathematics lessons in both schools to determine the usual proceedings and prepare the replacement of parts of the lessons with writing-to-learn tasks in the experimental condition. In both schools, teachers first explained new theory from the textbook about the topic at stake, after which students carried out assignments related

to the explained theory. For homework, students completed assignments if they failed to complete them during the lessons.

In preparation for the intervention, the first author cooperated with the teacher students to design three writing tasks, accompanying model texts, and a teacher's manual. First, the first author and the teacher students determined collaboratively which lesson series from the textbook was most suitable for embedding writing tasks (both schools used the same textbook). They discussed which parts of the business-as-usual lessons to replace with writing tasks. Second, the first author proposed topics for the writing tasks and model texts. The teacher students decided whether these topics involved important aims of the lesson series. Finally, the teacher's manual composed by the first author was checked by the teacher students to ensure that suggestions for explaining conceptual relations in the model text and for lesson proceedings were clear.

For composing the writing tasks, it was decided to use the conceptual relations 'condition', 'sequence', and 'comparing' because these were most important in the business-as-usual lessons. These conceptual relations match the genre explanation (Rose, 2008). Therefore, this genre was used in all three writing tasks. The writing tasks required students to write an explanation of how mathematical tasks can be carried out, each focusing on a different conceptual relation. Additionally, each writing task required students to write their explanation for an audience consisting of sixth grade students who have no prior knowledge about the mathematical tasks described. To enable their audience to understand the text, students needed to simplify their formulations thoroughly.

GWPR instruction entails that each writing task is preceded by a teacher's explanation of a model text. The model texts were based on texts from the students' mathematics textbook and were rewritten in such a way that they were comprehensible to an audience of sixth-grade students. The topics of the model texts were related to the topics of the writing tasks in such a way that students could use the text as an example for writing. To avoid copying, the topics of the model texts and aligned writing tasks were not identical. For instance, the topic of a model text was 'which shop is the cheapest when taking account of shipping costs and the price?' while the topic of the aligned writing task was 'where can you buy the most seeds for a specific amount of money?' Each model text contained one conceptual relation that was expressed in various ways. For instance, the relation 'condition' was expressed as: 'if you know … you can …'; 'if you make … you not only give... but you …'. The purpose was to familiarize students with various realizations of a conceptual relation. As is standard in texts about mathematics (Veel, 1999), the model texts contained graphical representations, such as a table, in addition to the text. The different stages of GWPR instruction on writing-to-learn (planning, individual formulation, review, and final draft) were the same as in Study 1.

4.7.4 Instruments

4.7.4.1 Prior knowledge tests

In consultation with the two teacher students, the first author composed the tests for determining students' prior topic knowledge and insight. The test was based upon test items from the textbooks students had used in grade 9 (Reichard, 2011a; Van Ber-

chum, 2010) and in grade 10 in the period preceding the experiment (Reichard, 2011b). The items concerned mathematical skills, such as understanding and drawing graphs, and elements of the comprehensive theme 'mathematical relations', for instance 'equations'. Prior topic knowledge was measured using 10 multiple choice items; prior insight was measured using five open-ended questions (with fixed answers).

The homogeneity of the tests measuring prior topic knowledge and insight is low (Cronbach's alpha is −0.14 and 0.15 respectively), which was expected because of the large variation in mathematical subjects tested. Cronbach's alpha provides an underestimation of test reliability (Boyle, 1991; Sijtsma, 2009; Taber, 2018). Therefore, the tests might still be sufficiently reliable to explain variance in our posttest measures. Therefore, we decided to include both tests as covariates in our analysis.

4.7.4.2 Vocabulary

A vocabulary test of 30 items derived from the Dutch version of the Peabody Picture Vocabulary Test was composed. For the selection of words, we used four sets (nr. 9, 10, 11, 12) aimed at all ages above 15. Because the students belonged to the youngest in this category, we selected words with p-values on or above 0.80. Cronbach's alpha was 0.89.

4.7.4.3 Topic knowledge and insight in the post-test

Unfortunately, topic knowledge could not be measured because of test regulations in one school that excluded multiple choice questions from mathematics tests. Insight was measured by means of a post-test consisting of 17 open-ended questions, which were mathematic sums. These questions were selected in consultation with the teachers from two final tests completing the theme 'mathematical relations' in the used textbooks.

Scores on the post-test were based on one or more elements students' answers should comprise. For each element, one point was assigned. The number of elements varied per item. The maximum score for the 17 items was 33 points. Cronbach's alpha of the open-ended questions was 0.78. Given that the items of the post-test were derived from the narrow domain of mathematical relations that was taught in the lessons (unlike the prior knowledge tests), a rather high homogeneity was expected. Inter-rater reliability was not computed because the answers on the open-ended questions (sums) were fixed.

4.7.3.4 Post-test writing task

Finally, a post-test writing task was administered. The task consisted of two parts: a computation and a writing task. First, students had to make the computation and next the writing task, which was writing an explanation of how they had carried out the computation in the first part. Students were asked to write this explanation for an audience of sixth-grade students. This writing task was used for examining the process of writing-to-learn in the think aloud study.

4.7.4 Procedure

Mathematics lessons were scheduled twice a week, with a lesson duration of 45 min. In total, nine lessons were involved in this study. The first was used for administration of the tests for prior insight, topic knowledge, and vocabulary. The second, fourth, and sixth lesson were used for the first part of each writing task (writing the first draft). The third, fifth, and seventh lesson were used for the second part of the writing task (writing the final text). In the eighth lesson, the post-test for insight was administered, and in the ninth lesson, students performed the post-test writing task.

Table 4.8 presents the lesson structures for the control and experimental conditions. The lessons in the control condition consisted of the teacher's explanation of new theory, students' completing assignments from the mathematics textbook, and completing unfinished assignments for homework. To keep time on task equal in both conditions, the main elements of the business-as-usual lessons were replaced by a writing task in the experimental condition. Each of the three experimental writing tasks were performed in two lessons. Explanation of the model text, planning, and writing the first draft (the first part) took place in one lesson, and reviewing and revising the draft (the second part) in the subsequent lesson. Teachers' explanations of new theory were replaced by homework requiring students to study new theory from their textbooks. Because students were not used to studying theory individually, they received instruction on how to perform their homework in lesson 1.

For securing treatment fidelity, the first author observed whether the teacher students carried out the lessons as intended and as described in the teacher's manual in all experimental lessons. No deviations in delivery of the experimental condition were encountered. However, in one school, a change of scheduling because of unexpected staff meetings became necessary. Therefore, the two parts of the second and third writing tasks had to be merged into one lesson each. This resulted in the loss of two (of six) lessons from the series for the two classrooms involved.

Table 4.8 The lesson structure in control and experimental conditions in mathematics class

Control group	**Experimental group**	
Business as usual (lessons 2-7)	**First draft (lessons 2, 4, 6)**	**Revision (lessons 3, 5, 7)**
• Explanation of new theory • Performing assignments from the textbook	• *Writing to learn task* part 1: explanation of model text, planning and writing a first draft	• *Writing to learn task part* 2: feedback and writing final text
Homework: completing unfinished assignments		Homework: studying new theory

Eight students (four experimental and four control) were randomly selected and performed the writing task thinking aloud. The think aloud sessions took place the day after the post-test had been administered. Students performed the writing task individually in an empty classroom, supervised by the first author. She provided them

with the same type of instruction as in Study 1 (see Procedure). Students wrote their texts by hand. Their performance was video recorded. The duration of the sessions was 15–20 min.

4.7.4.1 Coding students' transcribed utterances

The coding scheme consisting of 29 codes was the same as the one used in study 1 (see Appendix A). By means of this scheme, a researcher and a research assistant coded the utterances, one code per utterance or per (non-verbal) instance. For determining inter-rater reliability, two protocols (one for each condition) were coded independently by two raters. There was agreement for 78% of all utterances and instances. We consider this a sufficient reliability of coding. Differences in coding were resolved after discusion.

4.7.4.2 Data analysis

ANOVA was used for comparing students' prior insight, prior topic knowledge, and vocabulary in the two conditions. By means of ANCOVA, differences between students' post-test scores on insight in the two conditions were compared. Prior insight, prior topic knowledge, and pre-test vocabulary were used as covariates. In all statistical tests, alpha level was 0.05. The analysis of the process of writing-to-learn was performed as described in Study 1.

4.7.5 Results

Table 4.9 presents the means and standard deviations for four variables involved: prior insight, prior topic knowledge, vocabulary, and post-test insight. Scores on the latter test were based on one or more elements students' answer should comprise. For each element, one point was assigned. The number of elements varied per item. The maximum score for the 17 items was 33 points.

Table 4.9 Means and standard deviations for prior knowledge (insight), prior topic knowledge, vocabulary knowledge, and post-test scores on insight

Variables	Experimental group ((N=30) Mean (SD)	Control group (N=37) Mean (SD)
[1)]Prior insight	.60 (.62)	.65 (.72)
[2)]Prior topic knowledge	5.00 (1.46)	4.46 (1.33)
[3)]Vocabulary	21.67 (5.03)	22.05 (5.05)
[4)] Post-test insight	19.67 (6.23)	12.95 (7.51)

[1)] theoretical maximum score: 5 [2)] theoretical maximum score: 10 [3)] theoretical maximum score: 30
[4)] theoretical maximum score: 33

4.7.5.1 Differences in prior insight, prior knowledge, and vocabulary

ANOVA showed that experimental and control students' scores on prior insight and prior knowledge did not differ significantly: respectively, $F(1,65) = 0.086$, $p = 0.77$, partial $\eta^2 = 0.001$, and $F(1,65) = 2.51$, $p = 0.118$, partial $\eta^2 = 0.037$. No significant differences in vocabulary knowledge between the two conditions were found either: $F(1,65) = 0.098$, $p = 0.76$, partial $\eta^2 = 0.002$.

4.7.5.2 Differences in post-test insight

An ANCOVA showed that the covariates vocabulary and prior topic knowledge were not significantly related to post-test insight: respectively, $F(1,65) = 1.77$, $p = 0.19$, partial $\eta^2 = 0.03$, and $F(1,65) = 1.13$, $p = 0.29$, partial $\eta^2 = 0.02$. The covariate prior insight predicted the scores on post-test's open-ended questions significantly: $F(1.65) = 9.48$, $p = 0.003$, partial $\eta^2 = 0.13$. Therefore, in the final analysis, only the covariate prior insight was included. This showed that the control and experimental group differed significantly in post-test insight. The experimental group scored higher than the control group: $F(1,65) = 18.84$, $p = 0.00$, partial $\eta^2 = 0.23$. Additionally, the covariate prior insight was related significantly with the dependent variable: $F(1,65) = 8.98$, $p = 0.004$, partial $\eta^2 = 0.12$.

4.7.5.3 The process of writing-to-learn

Table 4.10 shows the proportioned means and standard deviations for each of the seven selected indicators of writing-to-learn for experimental and control students and the effect size of the difference. The results of all codes in the think aloud study are given in Appendix D.

Table 4.10 Proportioned means and standard deviations and effect sizes of indicators of writing-to-learn in experimental group (N=4) and control group (N=4)

	Experimental	**Control**	**Effect size**
Codes	Mean (SD)	Mean (SD)	Cohen's d
PLANNING: generating			
Using knowledge about audience	.03(.05)	.01(.02)	.53
PLANNING: selecting			
Thinking about content selection	.00 (.00)	.01 (.05)	-.29
FORMULATING			
Thinking about formulating	.00 (.00)	.00 (.00)	.00
Revising while formulating	.00 (.00)	.00 (.00)	.00
Revising after finishing an utterance	.02 (.02)	.00 (.00)	1.43
MONITORING			
Rereading own text	.07 (.13)	.03 (.03)	0.42
Rethinking task approach	.16 (.19)	.10 (.09)	0.40

Table 4.10 shows differences in favor of the experimental students on the following reflective activities: using knowledge about audience, d = 0.53 (medium effect); revising after finishing an utterance, d = 1.41 (large effect); rereading own text, d = 0.42 (small effect); and rethinking task approach, d = 0.40 (small effect). Two reflective activities, thinking about formulating and revising while formulating, were not used in both groups. Thinking about content selection took place in the control group only (small effect).

4.7.6 Discussion

We assume that writing can be a tool for learning that is suitable not only for high-achieving students but also for low achievers. When students write down their thoughts, they can read and reread these externalized thoughts and reflect on their intended meaning. In doing so, they may start questioning rhetorical issues, such as the comprehensibility of their ideas to the intended readers. In turn, this may lead to new insights and the acquisition of new topic knowledge.

GWPR instruction stimulates students' reflection repeatedly. For low achievers, a strength of GWPR is that reflection on planning and reviewing activities takes place in peer discussions that support them in learning (Rivard, 2004). Additionally, it provides students with the opportunity to discuss their thoughts using everyday language instead of academic language, which may stimulate low achievers' learning as well (Rivard, 2004; Akkus et al., 2007). We expected that GWPR instruction would lead to new insights for low achievers, which was the case for mathematics, but not for biology. This outcome may be explained by the age difference between the biology (12–13) and mathematics students (15–16) in our studies. The younger students may have had more difficulty with reflecting on their texts. Although studies conducted with young students (grade 6) led to positive effects of writing-to-learn (Ritchie et al., 2011; Martinez et al., 2015), these studies were not specifically directed at low achievers.

However, a more plausible explanation for the disappointing result in the biology classes is that time on task plays an important role in learning by writing. In mathematics class, the two parts of the writing tasks, 'planning and writing a draft' and 'reviewing and revising the draft', were spread over two 45 min lessons. In biology class, however, both parts of the writing tasks had to be carried out in one lesson of 45 min. This may have been too short for the students to really experience the importance of the different steps to be taken in GWPR instruction. Probably, the explanation of the conceptual relation at stake in each writing task and the peer discussions on planning and reviewing (feedback) require more time to be grasped than was available in one single lesson. The review phase in GWPR (including feedback from peers and text revision based on that feedback) is something that students are not especially used to do in regular education (Kiuhara, Graham & Hawken, 2009; Van Gelderen & Blok, 1991). Therefore, they may have completed their writing tasks rushing without understanding the purpose of revising. In addition, an advantage of splitting each writing task in two lessons is that students have the opportunity to view their drafts with 'new eyes' because their memory of writing the first draft has somewhat faded. This provides a better motivation for them to take the revision of their texts more seriously.

Hand et al. (2021) reviewed studies on the Science Writing Heuristic (SWH), a tool for learning by writing a lab report. They concluded that the time on task is an important factor for student learning. Their explanation is that students need sufficient time for becoming familiar with new ways of learning. This probably applies to GWPR instruction as well. It may be especially true for low achievers because they need more time for learning in general than average or high achievers do.

4.7.6.1 The writing-to-learn process

The assumption behind the think aloud studies was that GWPR instruction stimulates a recursive process between content and rhetorical goals (Bereiter & Scardamalia, 1987; Galbraith, 2009) taking place during planning and revising activities. Seven reflective activities were labeled as indicators of this recursive process (see Table 11). The outcome of the think aloud studies seems to echo the results of the quasi experiments. Experimental students in biology differed very little from control students in indicators for writing-to-learn (one positive and one negative effect), whereas experimental students in math differed substantially from control students on four indicators.

A remarkable difference found between the experimental biology and mathematics students is that the first frequently reflected on content selection (a large effect), whereas mathematics students did not (a small negative effect). This difference can be explained by the nature of the writing tasks. Biology students' writing task required them to use their knowledge of the musculoskeletal system for explaining how to act in a given situation. Therefore, students needed to select which parts of the system to use in their writing. Mathematics students, however, were instructed to describe how they solved a sum. They did not need to reflect on content selection because the content (the computation of the sum) was already there. This demonstrates that reflective activities analyzed in think aloud studies are dependent on the specific demands that each writing task poses. One should not expect the same types of reflective activities to emerge for different writing tasks unless the cognitive demands of these tasks are similar.

4.7.6.2 Students' and Teachers' Evaluation

To explore students' and teachers' view on learning by writing, we asked them how they valued the writing-to-learn lessons. How did they experience this new way of learning? Did it support students? These questions were posed in interviews with three students in biology class, four students in mathematics class, and the teacher students that had instructed the experimental classes. Furthermore, 31 and 28 students respectively from the experimental biology and mathematics classes filled in a questionnaire containing four multiple choice questions.

Reactions by the students in biology class were mixed. Two (of three) interviewed experimental students found the writing-to-learn tasks quite difficult. One of them kept struggling during the experimental lessons because he did not understand what to do. The third student stated that the writing had improved her understanding of the biology subject matter involved. In their answers on the questionnaire, one (of 31) student found the writing to learn tasks difficult, 14 (of 31) students found the writing tasks a little complicated, whereas 16 students considered the tasks easy. A

number of ten students answered they had learned from executing the writing tasks, 14 students had learned a little, and seven indicated they had not learned anything. Both biology teacher students stated that they viewed writing a good way to stimulate students' learning. In class, one teacher noticed that some students acquired new insights. However, he also had observed that students had to hurry to finish the writing tasks. He suggested planning a longer lesson duration.

The mathematics students who were interviewed stated that it does not matter which type of learning is used. However, writing took more effort. Writing a draft required deep thinking about organizing, as two students reported. One student who was interviewed remarked that giving feedback to her peer provided her with ideas just as her own writing. Ten (of 28) students answering the questionnaire found the writing-to-learn tasks difficult, 11 students found it a little complicated, and only seven students had no problem with them. However, 15 students believed they had learned from the execution of the writing tasks, while 11 students believed they had learned a little. Two students thought they did not learn anything.

The two mathematics teacher students considered writing-to-learn an effective way of learning. One teacher observed that students worked intensively when carrying out the writing tasks. She stated that more time for using writing-to-learn tasks might have made students feel more comfortable. The second mathematics teacher considered writing-to-learn a valuable addition for teaching mathematics because it is a different way of thinking than usual in mathematics.

In summary, students' reactions reflect the outcomes of the two studies. Their teachers took part in the study because they were enthusiastic about writing-to-learn and remained so. They attributed students' discomfort to a lack of time for executing the writing tasks.

4.7.7 Suggestions for future research

As previously discussed, research investigating effects of instruction in writing-to-learn specifically directed at low-achieving students is scarce. Nevertheless, there are good reasons to expect that low-achieving students can benefit as much from instruction in writing-to-learn as higher-achieving students (Akkus, Gunel & Hand, 2007; Faber, Morris & Lieberman, 2000; Rivard, 2004). Our studies give further support to this view. Although the biology study did not lead to the expected results on low achievers' learning, it certainly provided valuable information about the conditions that must be fulfilled for an elaborated approach such as GWPR to become successful. Therefore, future research allowing low achievers sufficient time on task in familiarizing themselves with the different steps of GWPR (such as recognizing genre specific conceptual relations, using them in writing tasks, and peer discussions directed at planning and revising a draft) is highly recommended. This research is essential to test the viability of assumptions underlying instruction in writing-to-learn and more specifically GWPR instruction for low-achieving students in diverse educational contexts, disciplines, and ages. The present research base needs more systematic trials to support the positive expectations of the effects of writing-to-learn for low-achieving students.

This type of future research can be strengthened by paying particular attention to the sample size, especially when the focus is on the processes of writing-to-learn. In our studies, there were only small samples selected for these analyses

because of the workload involved in the analysis of think aloud protocols. In future studies, it is important that the generalizability of our findings regarding the indicators of writing-to-learn is tested with larger samples. In addition, other methods for tracking low achieving students' writing processes such as keystroke logging (Baaijen & Galbraith, 2018) and eye movement tracking could be used to find additional marks of students' reflective processes while writing.

Another important issue for future research is randomization on the student level. While our studies were quasi-experimental in using intact classrooms for assignment to an experimental or control condition, a more profound design for arriving at causal interpretations of effects found is desirable. Although it is difficult to assign students in the same classroom to different conditions because of 'leakage' of the instructional components, there is a way of dealing with that problem. Care should be taken to make the materials self-instructing (such as the analysis of conceptual relations in model texts) and the lesson structure identical for both conditions. Additionally, the topics for the writing tasks should systematically differ between the conditions by using different overarching themes from different lesson series in the discipline (e.g., 'the human body' vs. 'bacteria' in biology). Effects can be analyzed by comparing topic knowledge and insight on the two different themes that students in the two conditions in the previous lessons had been writing about.

4.7.8 Pedagogical implications

Regular mathematics lessons often consist of the teacher's explanation of new theory and thereafter students' completion of mathematics tasks. Students are supposed to learn by listening to the teacher and applying the explained theory by carrying out tasks from their textbooks. Low-achieving students may have trouble asking questions in response to the teacher's explanation because it can be difficult for them to formulate what they do not understand. This can result in their inability to carry out mathematics tasks.

Results of our studies show that low achievers benefit from writing-to-learn in mathematics as measured by their insight in how to execute mathematical tasks. Additionally, it seems that students receiving GWPR instruction show more indicators of reflective activities in a post-test writing task. Therefore, we assume that GWPR instruction for executing writing tasks allows low-achieving students to become more actively engaged in learning new theory. This engagement consists of discussions with their peers for planning, feedback, and revising a first draft directed at comprehensibility for an audience with no prior knowledge of the mathematical tasks involved. However, teachers may be hesitant to use writing in their math class because they believe that evaluating students' texts increases their workload (Countryman, 1992). In contrast, it must be emphasized that explicit evaluation by the teacher of students' written texts is not necessary because writing is aimed at reflecting on and learning of mathematics. In addition, the advantage of writing-to-learn tasks is that reading students' texts can inform about students' thinking, which may help teachers to focus on students' needs (Countryman, 1992).

Teachers in the content areas (such as biology or math) do not seem at ease with developing writing-to-learn tasks for their classrooms because they are not used

to writing as a tool for learning. As our studies show, writing tasks can quite easily be embedded in regular education as alternative for the more traditional approach of teaching. A more difficult aspect for content area teachers in GWPR instruction is the construction and use of model texts to demonstrate different ways of formulating genre specific conceptual relations. Implementation of that part of GWPR in classroom practice therefore requires the support of language teachers and/or the provision of a rich set of examples demonstrating the use and explanation of conceptual relations in context. Preferably, this type of support is offered on school level as part of teachers' professional development program.

4.7.9 Conclusions

The present study was directed at writing as a tool for learning for low-achieving students in secondary education. We investigated whether GWPR instruction (Genre Writing with Planning and Revision) leads to positive effects on these students' learning. Firstly, we investigated whether GWPR instruction focusing on explicit genre knowledge, cognitive and metacognitive strategies, and writing for a lay audience leads to more insight and topic knowledge in biology and mathematics classes compared to business-as-usual lessons (not including writing activities). Secondly, we investigated whether students that had received GWPR instruction showed more signs of the process of writing-to-learn, such as reflection directed at generating, selecting and organizing text contents, and reviewing and revising formulations.

Results showed that GWPR instruction for low-achieving biology students did not lead to effects on topic knowledge and insight, while GWPR instruction for low-achieving mathematics students did lead to a large effect on insight. Unfortunately, in mathematics classes, topic knowledge could not be measured because of school regulations prohibiting multiple choice questions in math examinations.

Think aloud studies were carried out with a random selection of students to identify the process of writing-to-learn in both disciplines. Based on previous studies, reflective activities for planning, formulating, and monitoring are regarded as indicators of the process of writing-to-learn. Seven such indicators were identified. We investigated whether GWPR instruction had incited the process of writing-to-learn in a post-test writing task for experimental students in comparison to control students.

Table 4.11 summarizes the results for both studies. For biology, analysis of the think aloud protocols showed that the activity thinking about content selection was used more frequently by experimental than by control students (large effect), while the activity revising while formulating was used more often by the control students (large effect). For mathematics, the following indicators were more frequently used by the experimental than the control students: using knowledge about audience (medium effect), revising after finishing an utterance (large effect), rereading own text (small effect), and rethinking task approach (small effect). The indicator thinking about content selection was used more often by the control students (small effect).

Thus, in the mathematics study, experimental students showed several indications for writing-to-learn, while in the biology study they showed only one effect in favor of the experimental students and one opposite large effect in favor of

the control students. These results are in line with the outcomes of the quasi experiments in the studies 1 and 2, showing no effect of GWPR instruction on learning for biology and positive effects on insight for mathematics.

Table 4.11 Effect sizes of indicators for learning by writing in study 1 (biology) and study 2 (mathematics)

Indicators for learning by writing	Biology Cohen's d	Mathematics Cohen's d
Planning		
Using knowledge about audience	Not used	Medium
Thinking about content selection	Large	Small negative effect
Formulating		
Thinking about formulating	No effect	Not used
Revising while formulating	Large negative effect	Not used
Revising after finishing an utterance	No effect	Large
Monitoring		
Rereading own text	No effect	Small
Rethinking task approach	No effect	Small

Not used: not used by both conditions

Chapter 5 General discussion

Chapter 5

General discussion

5.1 Introduction

Writing-to-learn requires students to carry out writing tasks about a disciplinary subject. Writing-to-learn tasks are intended to stimulate reflection on the disciplinary topic students write about, because this may lead to learning (Bereiter & Scardamalia, 1987). An example of a writing-to-learn task is explaining a subject adapted to the needs of readers who have no prior knowledge of it. The task requires writers to 'translate' academic language into a text which is comprehensible to the target audience by using everyday language. The assumption is that students acquire more insight into a topic by reflecting on how to explain subject matter to an uninformed audience.

We consider writing-to-learn a promising way of learning, additional to the more usual ways such as listening to teachers' explanation of subject matter followed by class discussion. The advantage of writing-to-learn is that students record their thoughts on paper, offering them the opportunity to read and reread their drafts and reflect on them. In doing so, students can use their writing as an external log of their thinking and reduce the appeal on their working memory.

Concerning the question how to use writing-to-learn in class, a first assumption was that all writing leads to learning (the strong text theory, see Chapter 2). However, Bangert-Drowns et al. (2004) suggested that instruction in cognitive and metacognitive strategies should accompany the writing process for effects on learning. Following this line of thought, many studies using various types of instruction on such strategies were undertaken. Reviews confirm that writing combined with instruction directed at different cognitive and metacognitive strategies can lead to learning (Miller et al., 2018, Graham et al. 2020, Hand et al., 2021). This result has been found for different disciplines and different levels of education. However, these reviews did not succeed in defining which specific types of instruction for writing-to-learn were more successful than others (Graham et al. 2020). The present research aimed at investigating which types of instruction lead to effects on learning and in which way.

In this thesis, we defined 'type of instruction' as instruction directed at one or more cognitive or metacognitive processes intended to stimulate the process of writing-to-learn. Instruction may take the form of, for instance, planning the organization of the text, revising the text in view of the needs of the audience or increasing the students' knowledge of the genre in which they are supposed to write. Nine studies were conducted: a review study, four quasi-experimental studies and four think aloud studies. The review study was intended to acquire an overview of different types of instruction in research on writing-to-learn and the effects on learning. This resulted in four types of instruction: Forward Search, Backward Search, Genre Writing and Planning Only. We will discuss these four types in more detail in the next section.

In the review study, we concluded that Genre Writing with Planning and Revising (GWPR), appeared to be the type of instruction that was most promising. Therefore, we investigated effects of GWPR-instruction by means of four quasi-experimental studies, two aiming at teacher students and two at low achieving students. All quasi-experimental studies contained a think aloud component directed at finding indications of participants' writing-to-learn processes.

In what follows, we discuss the main findings of the review study. Subsequently, we go into detail about GWPR-instruction, before we discuss the main findings of the quasi- experiments and the think aloud studies. Then, we provide explanations for the outcomes. Thereafter, we discuss limitations of the studies, suggestions for further research and finally implications for classroom practice.

5.2 Main findings

5.2.1 Review study

Chapter 2 comprises a review study that was undertaken to acquire insight into the types of instruction used in studies on writing-to-learn, and to investigate which types resulted in positive effects on learning. More specifically, we investigated which types of instruction led to positive effects on insight into the meaning of newly learned knowledge and topic knowledge. By *insight* we mean the ability to relate new concepts to prior (topic) knowledge. By *topic knowledge* we mean recall of learned concepts (retrieving concepts from memory). In this thesis, we take the view that both types of knowledge are relevant. We view *insight* as the ultimate aim of writing-to-learn, because insight is the manifestation of higher order understanding of concepts.

To acquire a broad overview of research into instruction in writing-to-learn, we selected 43 empirical studies conducted in many types of education varying from grade 5 to higher education, and in various disciplines. All selected studies were embedded in regular education, because we wanted to investigate how writing-to-learn operates in classroom contexts. Quantitative as well as qualitative studies were selected to get the best possible picture of the used modes of instruction. For determining which modes of instruction led to positive effects on learning, we used the studies comparing experimental with control groups.

In the selected studies, four types of instruction were distinguished. These are Forward Search (instruction on reviewing and revising a draft), Backward Search (instruction on planning, reviewing and revising a draft), Genre Writing (explanation of the genre students are required to write in, sometimes combined with instructions on planning or on planning and/or revising a draft), and Planning Only (instruction on planning a draft). The first three types were based on Klein's (1999) hypotheses about the process of writing-to-learn and the fourth emerged from the reviewed studies. After the identification of instruction in all 43 studies, we determined for each of the four types of instruction the percentages of experimental comparisons with positive effects on learning. The number of studies comparing experimental with control conditions was 24. The percentages were as follows: Forward Search: 67%, Genre Writing: 65%, Backward Search: 50%, Planning Only: 73%. The other 19 studies were studies comparing experimental conditions or case studies.

Compared to the other three types of instruction, Genre Writing including combinations with planning and revising (GWPR), appeared to be supported the most firmly by the empirical studies: 11 out of 14 studies showed positive effects. Therefore, in the next section we elaborate on this type of instruction only. The remaining types of instruction also had a fair success rate but were object of far fewer studies.

5.2.2 GWPR-instruction in the four quasi-experimental studies

Before we discuss the findings of the quasi experiments using GWPR-instruction, an explanation of how we used GWPR to stimulate learning follows.

We fine-tuned GWPR-instruction to the educational contexts in which we wanted to experiment. GWPR-instruction is based on the view that students need genre knowledge for understanding and writing in a genre. Therefore, we selected conceptual relation(s) characteristic of a genre and made these explicit in the writing-to-learn tasks to be used in the lessons. As an extra element, we added that students write for a lay audience in order to stimulate reflection.

GWPR-instruction was applied as follows. In the first place, genre knowledge was provided in a preparatory activity (before the writing assignment) by means of a model text of the genre at stake. Genre knowledge is defined as knowledge of the genre's rhetorical goal and prevalent conceptual relations between text elements to arrive at this goal (Halliday & Martin, 1993). For instance, the rhetorical goal of explanatory texts is clarifying a topic, and a prevalent conceptual relation, for instance 'comparing', shows the disciplinary use of the genre. Analyzing a model text can make students aware of these characteristics. The model text should therefore exemplify various linguistic realizations of the conceptual relation, for instance the relation 'comparing' is realized as 'X differs from Y'. If students are made aware of how the conceptual relation 'comparing' in an explanatory text can be realized, it is easier for them to make comparisons in their own writing and to reflect on the results, which may lead to new insights into the meaning of these comparisons (Langer & Applebee, 1987). In addition, in GWPR the model text is written in such a way that it is comprehensible to a lay audience, which means, for instance, that the model text does not include disciplinary jargon.

In the second place, planning consists of selecting and organizing content in preparation of writing. In our view of GWPR, instruction on planning entails that students can brainstorm in pairs to select knowledge elements that they find relevant. Students are instructed to represent their selection by means of keywords and to organize them in a mind map, such that the structure of their draft becomes visible. They thereby have to consider the conceptual relations in view of comprehensibility for their audience and may therefore decide to include an introduction or a conclusion. While writing their drafts they are supposed to consult their planning as well as the model text exemplifying the formulation of central conceptual relations.

In the third place, revising consists of the use of metacognitive strategies for reformulation on the basis of peer feedback. Students are instructed to review their peer's draft by focusing on the conceptual structure as realized in the text and on its comprehensibility for a lay audience. In this process, students reflect on their peer's representation of the conceptual relations in language. Additionally, they put themselves in the point of view of the intended (lay) audience. Students use peer

feedback for revising their drafts individually. In doing so, they have to reflect on their original insights in the conceptual relations and their original formulation, which may lead to new insights (Bereiter & Scardamalia, 1987).

By using GWPR-instruction in this way, we intended to offer all types of students as much support as possible in their efforts at writing-to-learn.

5.2.3 Two studies directed at biology and mathematics teacher students

In order to investigate to which extent GWPR-instruction leads to new insight and topic knowledge in teacher students, two studies were conducted: one directed at teacher students in biology, the other at teacher students in mathematics. We chose to investigate teacher students in these two disciplines, because they vary in the amount in which they rely on writing: in biology, writing is a regularly used activity, for instance for writing lab reports, whereas in mathematics writing is scarcely used.

The first study was directed at 38 third year biology teacher students, the second at 62 third year mathematics teacher students. A post-test-only design was used comparing an experimental to a control condition. The dependent variables were insight and topic knowledge. Prior knowledge of the discipline and vocabulary knowledge were used as covariates. While the experimental groups received the lessons including writing-to-learn tasks, the control groups received the regular lessons (business as usual). The interventions were embedded in regular classes, achieving an ecologically valid approach. The researcher and regular teacher co-created the model texts, writing tasks and teacher's manuals.

We found that GWPR-instruction led to positive effects on insight and topic knowledge for biology teacher students (large effect size), and on insight for mathematics teacher students (large effect size). Because the post-test for topic knowledge on mathematics was not reliable, results for topic knowledge could not be included in the analysis. These findings show that GWPR-instruction can be useful for teacher students in disciplines that differ largely in their regular use of in-class writing activities.

5.2.4 Two studies directed at low achieving students learning biology and mathematics

In order to investigate whether GWPR instruction is useful for low achieving students, one study was directed at students in biology, the other at students in mathematics. We investigated effects of GWPR-instruction in grade 7 biology class (n = 75) and grade 10 mathematics class (n = 67) of prevocational education. We reasoned that writing-to-learn benefits these students, because writing down their thoughts reduces their cognitive load, and therefore facilitates reflection on ideas and transforms them. Just as in the studies directed at teacher students, post-test only-designs were used. While the experimental groups received the lessons including writing-to-learn tasks, the control groups received the regular lessons (business as usual).

The structure of GWPR was the same as described for teacher students. The contents and size of model texts were adjusted to the grade levels of the students just as the writing tasks. In these studies, the teachers were teacher students in their fourth (and last) year, who had participated in the experimental condition of the above described studies directed at biology and mathematics teacher students. Therefore, they were familiar with writing-to-learn. The intervention was embedded in regular

classes, and the researcher and teachers co-created the materials, that is model texts, writing-to-learn tasks and teacher's manuals.

We found that GWPR-instruction did not lead to larger effects on insight and topic knowledge for experimental low achieving students learning biology when compared to control students. However, GWPR-instruction for low achieving students learning mathematics led to positive effects on insight. Topic knowledge could not be measured, because the schools did not allow multiple choice tests in mathematics. Based on these results, we concluded that low achieving students learning mathematics can profit from GWPR-instruction. We were, however, not able to show that low achieving students learning biology benefited from GWPR-instruction for learning.

5.2.5 Two think aloud studies directed at biology and mathematics teacher students

In order to identify indicators of the process of writing-to-learn, think aloud studies were conducted with the biology and mathematics teacher students. Experimental and control students carried out a post-test writing task while thinking aloud. This allowed us to compare experimental and control students' processes of writing-to-learn. We selected a random sample of the participants of the quasi-experiments (12 of 38 biology teacher students, 15 of 62 mathematics teacher students). The analyses of the think aloud protocols were directed at students' reflective activities, because these were considered indicators of the process of writing-to-learn (Bereiter & Scardamalia, 1987; Galbraith, 2009).

We analyzed the protocols of the think aloud sessions by means of seven codes, which we considered indicators of writing-to-learn: 'using knowledge about audience', 'thinking about content selection', 'thinking about formulating', 'revising after finishing an utterance', 'revising while formulating', 'rereading own text' and 'rethinking task approach'. It appeared that biology and mathematics teacher students who had received GWPR-instruction, respectively used three and four out of the seven hypothesized activities more often than the control groups. Two hypothesized activities, 'revising while formulating' and 'rereading own text', were undertaken more often by experimental biology and mathematics groups than by the control groups. A difference between biology and mathematics teacher students was that 'revising after finishing an utterance', was executed more often by experimental biology teacher students than by control biology teacher students, whereas it was used just as often by experimental mathematics teacher students as by control students. On the other hand, 'using knowledge about audience' was not used by biology teacher students, while mathematics experimental students used this indicator more than control students. Finally, 'thinking about content selection', was used more often by biology control students than by experimental students, whereas mathematics experimental students used it more often than control students.

The differences between experimental and control teacher students in favor of the experimental teacher students were relatively large in both disciplines. The differences between biology and mathematics indicators of writing-to-learn can largely be explained by differences in the demands of the respective writing tasks (see section 5.3). Differences found for the indicators of the writing-to-learn process in

this study largely confirmed advantages found for the two experimental groups on the post-tests measuring learning results.

5.2.6 Two think aloud studies directed at low achieving students' learning biology and mathematics

Similar to the teacher students, the low achieving (prevocational) students carried out think aloud writing tasks. Random samples of students were selected from all participating students in biology (10 out of 75) and mathematics (8 out of 67). The analyses of the think aloud protocols were performed in the same way as with the teacher students.

Experimental low achieving students learning biology appeared to show one indicator out of seven, 'thinking about content selection', more often than control students (large effect). Control students however performed one reflective activity more often than experimental students: 'revising while formulating' (large effect). Therefore, we were not able to conclude that experimental students in biology class showed more indicators of writing-to-learn than control students.

Experimental low achieving students learning mathematics showed four indicators ('using knowledge about audience', 'revising after finishing an utterance', 'rereading own text', 'rethinking task approach') more often than control students (medium effects). Therefore, we can conclude that experimental low achieving students having received GWPR-instruction in math class, show more signs of learning in their writing process than control students.

Overall, these findings from the two think aloud studies for low achieving students confirm the learning results found for experimental and control groups. Students learning biology did not show effects of GWPR-instruction for learning in their learning results nor in their writing process, but such effects were found for low achievers learning mathematics, both in their learning results and in the use of reflective indicators in their writing process.

5.3 Explanations

The following section 5.3.1 provides explanations for the effects of GWPR-instruction on teacher students' learning in the disciplines biology and mathematics, and on low achieving mathematics students' learning. Additionally, the absence of effects for low achieving biology students is explained. Then, in section 5.3.2, the outcomes of the think aloud studies for teacher students and low achieving students are discussed.

5.3.1 Effects of GWPR-instruction on learning outcomes for teacher students and low achieving students

In the present research, we assumed that the support provided by GWPR-instruction is useful for teacher students as well as for low achieving students. Results showed that GWPR-instruction led to positive effects on learning by teacher students in both biology and mathematics and by low achieving students in mathematics. We explain

these positive effects as follows. GWPR-instruction stimulates students' reflection systematically in all phases of the writing process. During planning, formulating, reviewing and revising, students are required to reflect on two foci, that is how to relate concepts and whether the sentence structure and wordings are comprehensible to the intended audience. The systematic repetition of these two aspects provides students with multiple opportunities to enrich their topic knowledge and acquire new insights.

However, experimental low achieving students learning biology did not show positive effects on learning. The lack of positive effects may be due to the fact that lesson duration was too short and too condensed for the students to become familiar with the process needed for formulating conceptual relations directed at a lay audience. The low achieving students learning biology had to carry out a complete writing-to-learn task in only one lesson (50 minutes). In the other studies (directed at teacher students and low achieving students learning mathematics) we divided each writing-to-learn task over two lessons. The study on low achievers learning biology was the only one taking place at the end of the school year. In that period, the school schedules did not allow us to organize planning and writing a draft in one lesson, while reviewing and revising a draft by using peer feedback, in a next lesson. The schools could not provide sufficient lessons for executing writing-to-learn tasks, because of activities planned, such as excursions. In this context, it is relevant that Hand et al. (2021) concluded that providing sufficient opportunities for writing-to-learn is an important factor for becoming familiar with the process of writing-to-learn.

5.3.2 The process of writing-to-learn of teacher students and low achieving students

Experimental biology and mathematics teacher students showed more indicators of writing-to-learn than control students. We concluded that the larger frequency of these indicators is evidence for the process of writing-to-learn incited by GWPR-instruction.

There were however differences between the indicators found to be more frequent in experimental biology and mathematics teacher students. Experimental mathematics teacher students did not show 'revising after finishing an utterance' more often than the control group, whereas experimental biology teacher students did. This difference may be due to the relative novelty of writing activities in mathematics class. Mathematics teacher students were not used to rereading and revising an utterance yet and wanted to finish their task quickly.

Another difference was that experimental and control biology teacher students did not show the indicator 'using knowledge about audience' at all. Additionally, experimental biology teacher students did not show 'thinking about content selection' more often than the control group, whereas experimental mathematics teacher students used both indicators more often than the control group. These differences may be explained by differences in the writing tasks. While mathematics teacher students were instructed to write to grade 9-10 students just as in the writing-to-learn tasks, the biology writing task, a post-test, was different, as students were not instructed to write to any audience. Therefore, biology teacher students did not need to reflect on comprehensibility for their audience. Additionally, instruction mentioned specific contents from the textbook biology teacher students

were required to write about, while mathematics teacher students had to decide themselves what was relevant to include in their writing. Thus, there was no need for biology teacher students to think about content selection, but only about how to formulate the text.

Regarding experimental low achieving students learning biology, we assume that they did not reflect more often than control students on their writing activities during the think aloud sessions, because time allowed for GWPR-instruction was too short to become familiar with the process of writing-to-learn. Experimental low achieving students learning mathematics disposed of twice as much instruction time for executing writing-to-learn tasks. It is plausible therefore, that effects of GWPR-instruction on the indicators of learning by writing for these students were far more apparent.

5.4 Limitations

An important limitation of our studies is that the sample sizes are quite small, especially in the case of the think aloud studies. This limitation means that it is unclear whether the findings are generalizable to larger populations of students. Various circumstances in the schools involved resulted into small sample sizes for our studies. In teacher education, the number of students following a specific course is limited. In the years we carried out our research, only one class was available. In prevocational education, we could include no more than two classrooms in our studies, because we had only one trained teacher to deliver the lessons.

The small samples in our studies with 75 participants as the largest sample, are not exceptional. Our review study comprises 25 (of 43) studies with a sample of 75 or less participants This may have to do with the intensity of the intervention in most studies and the fact that teachers need to be trained in using writing-to-learn. It appears, however, not impossible to study large samples, when the circumstances are favorable. For instance, the study by Nam, Choi and Hand (2011) is based on a sample of 345 participants from three secondary schools. It was possible to use such a large sample because the study took place in a situation where an identical national curriculum is required in all schools for secondary education. This enabled researchers to use the intervention in the teaching program of six experimental groups, while the control groups followed the usual program. The intervention was less intense, because researchers did not compose writing-to-learn tasks, but used a ready-made checklist, the Science Writing Heuristic (SWH). This gave room to researchers for training teachers in using the SWH by means of a series of three workshops.

5.5 Suggestions for future research

Based on the outcomes of our study we suggest three topics for future research. The first has to do with the design of future research into GWPR instruction. The second is research into the three less explored types of instruction in our review study (Forward Search, Backward Search and Planning Only). In the third place, the use of keystroke logging is suggested in combination with think aloud protocols for research into the process of writing-to-learn.

First, our studies into the effects of GWPR-instruction were quasi-experimental. Intact classrooms were assigned to an experimental or control condition. Obviously, a true experimental design is preferred for arriving at causal explanations. However, the circumstances in the institutions involved forced us to use quasi-experimental methods. A more solid design for arriving at causal interpretations of effects found is desirable. For instance, researchers may randomly assign students in classrooms to two different conditions. For both conditions, the teacher may provide instruction about two different topics for writing-to-learn tasks, but students in one condition complete writing tasks for one topic, while students in the other condition complete writing tasks about the other topic. In this way, the two groups of students can function as each other's control group. The fact that students in both conditions are working in the same classroom but on different writing tasks is not necessarily problematic, because the lesson structure is identical for students in both conditions and the teacher can devote attention to instruction for both groups.

Second, in our review study, we found empirical evidence for the effectiveness of four types of instruction in writing-to-learn. These results, however, must be considered with caution, because they are based on small numbers of (quasi)experimental studies (Forward Search: 3, Backward Search: 5, Planning Only: 5). The results of Genre Writing instruction are however based on a fair number of 11 studies comparing experimental to control groups. It is possible that a particular type of instruction may lead to positive effects only in certain contexts, depending on the type of students. For instance, some writers may acquire more insights when instructed by means of Forward Search than by means of Backward Search, because they tend to start writing spontaneously without planning first (Kieft et al., 2007). Other students may show more learning when instructed by means of Backward Search. They may prefer to reflect on planning their text before writing (Kieft, 2008). In order to gain more certainty about the question which type of instruction in writing-to-learn is effective in which contexts or disciplines, it is important that experimental studies are directed to these issues in future research. The results may facilitate teachers in choosing the type of instruction that suits their situation and learning objectives.

Third, although the think aloud studies on the process of writing-to-learn largely confirm the outcomes of the quasi-experimental studies, there is uncertainty about the question whether students have put into words all reflective activities that could be considered indicators of the writing-to-learn process. It may also be that they had difficulty with wording some of their thoughts. Additionally, there is uncertainty about the generalizability of the think aloud analyzes, because the samples are small (varying between 8 and 15). Therefore, we suggest the use of keystroke logging for

future studies into the writing-to-learn process, as has been done by Baaijen (2012). The advantage of keystroke logging is that this method automatically captures all participants' keystrokes during writing. Baaijen (2012) proposed a procedure how to analyze keystrokes for determining indicators. She reports that evidence for writing-to-learn processes was found. At the same time, keystroke logging allows to analyze larger samples than possible in think aloud studies. Baaijen (2012) used 80 participants, which is a more solid basis for drawing conclusions than the small numbers of participants in our studies who were subjected to thinking aloud.

Contrary to our approach in the think aloud studies, Baaijen considers pauses of minimally two seconds as indicators of reflection. She uses the location of pauses for determining which type of reflection took place. For instance, a pause before a 'burst' (uninterrupted text fragment) is considered to be a reflection on a following action, such as planning. A combination of keystroke logging data and think aloud protocols for identifying indicators of learning by writing may be a valuable addition to findings reported in the existing literature.

5.6. Pedagogical implications

In our studies, it appeared that teachers need support before they start to use writing-to-learn in the classroom. Therefore, an in-service training is proposed in section 5.6.1. Next, in section 5.6.2, we describe what it takes to prepare for GWPR instruction in writing-to-learn in regular education. Finally, we discuss in section 5.6.3 the possibility to connect writing-to-learn to the general language policy of schools.

5.6.1 In-service training to familiarize teachers with writing-to-learn

In our quasi-experiments, we observed that teachers found it difficult to include writing-to-learn tasks in their lessons. In addition, they felt uncomfortable explaining conceptual relations in a model text. Possibly, biology and math teachers do not consider themselves competent in explaining such language related issues, as Baker (2008) reported from interviews with science teachers. Other teachers may have problems using writing-to-learn as well. Next to problems with clarifying conceptual relations in model texts, they may have trouble to do justice to the different phases in writing, such as brainstorming and organizing contents, reviewing and revising drafts, as with activities with the focus on specific conceptual relations and a (lay) audience. Teachers need to understand the function of conceptual relations as well as the reason why students are required to write to a lay audience. Besides that, coaching students in carrying out writing-to-learn tasks is not easy for teachers who are not accustomed to pay attention to the quality of students' writing.

The purpose of an in-service training taught by experts in writing-to-learn is to offer teachers the opportunity to get over their hesitations and become familiar with composing writing-to-learn tasks and using them. Topics for such a training are: characteristics of the process of writing-to-learn, how to stimulate the process by means of instruction, how to compose model texts focusing on specific types of

conceptual relations, how to prepare teachers for coaching students' writing process and how to instruct students to give feedback to each other when working in pairs.

The training can be organized by institutes for teacher training. Essential is that experts let teachers go through the different steps of the implementation process to arrive at teaching writing-to-learn in the classroom (see 5.6.2). The expert can require teachers of a same discipline to cooperate in small groups to prepare the use of writing-to-learn in the classroom. Cooperation gives teachers confidence in implementing the new way of learning (Bean, 2011). When writing-to-learn tasks are used in the classroom, teachers can make observations of each other's approach and discuss it afterwards in a training session. In doing so, the in-service training should clarify which problems teachers encounter and support them to solve these problems.

5.6.2 How to use GWPR instruction in writing-to-learn

In the following, we describe what it takes to use GWPR instruction in writing-to-learn in regular education. In our quasi-experimental studies, we cooperated with teachers in designing writing-to-learn instruction. On the basis of these experiences, we elaborate on the preparation needed for successful GWPR instruction in writing-to-learn. We describe six activities that can be executed by teachers of all types of disciplines from philosophy to mathematics.

The first activity is studying the curriculum for determining in which course and which lessons writing-to-learn tasks can be inserted. Courses that fulfill this requirement are directed at theoretical knowledge (as opposed to practical exercises), for instance a series of lessons in biology about 'firmness and movement' discussing the function of human muscles and bones. In addition, objectives of the course need to comprise insight and topic knowledge, such as 'insight into the structure of the human body' and 'knowledge of the names of human bones'.

The second activity is determining the topics for the writing-to-learn tasks that will be used. Prior to each lesson, students study chapter(s) from the textbook, for instance about the history of biology in a course on Biology and Science for biology teacher students. The teacher derives topics for the writing-to-learn tasks from the studied chapters. An example of a topic is 'the treatment of flu in the seventeenth century'. Because writing-to-learn tasks need to be inserted in a curriculum, the number of chosen tasks depends on the number of available lessons. For instance, for a lesson series of six weeks we suggest minimally three writing-to-learn tasks. The completion of each task should be spread over two lessons because students need time to become familiar with the various phases of learning by writing. We base this recommendation on the fact that in one of our studies (biology in first grade prevocational education) completing the task in one 50-minute lesson proved to be unsatisfactory (see Chapter 4).

As a third activity, genres and conceptual relations are chosen for the writing-to-learn tasks and model texts. A text genre like 'argument', can be characterized by its rhetorical goal, which is 'convincing'. Preferably, genres frequently used in the discipline are chosen, for instance 'opinion paper' in the discipline Dutch language requiring students to write opinion papers. The same applies to the choice of conceptual relations. These need to be characteristic for the discipline and applicable in the chosen genre, for instance 'statement and argument' for an opinion paper on the characteristics of the literary movement 'modernism'.

The fourth activity is designing model texts. A model text serves as an example of the type of text students are required to write. It is written in the genre students are supposed to write and exemplifying the conceptual relation that is focused upon in the writing-to-learn task. Because the model texts are meant to support students in executing the writing tasks, teachers have to compose the text for the intended (lay) audience. Teachers can use existing textbooks about the subject, aimed at the intended audience as sources for their writing and combine and rewrite fragments (see the example in Appendix E). Teachers make sure that the chosen conceptual relation is dominant in the text and presented in various formulations, for instance: 'compared to', 'contrary to', 'other than' as formulations for the conceptual relation 'comparing'. This frequent use is meant to direct students' attention to the conceptual relation and to show several ways to realize it in language. The topic of a model text needs to be different but related to the topic of the writing-to-learn task, such as 'using muscles for throwing a basketball' and 'using muscles for kicking a soccer ball' in biology. This is important to prevent students copying the model text.

The fifth activity is composing writing-to-learn tasks. By composing writing-to-learn tasks, the teacher creates an instrument that stimulates deepening students' learning of a studied topic. The tasks require students to write texts about the topics derived from their homework, in the chosen genre, and to use specific conceptual relations chosen by the teacher. Additionally, students must direct their texts to an audience without prior knowledge about the topic.

The sixth activity is organizing instruction for the different phases of the writing tasks. Teachers have to provide instruction in planning, reviewing and revising. They instruct students to brainstorm in pairs to formulate ideas for their texts, taking into account the conceptual relation and the (lay) audience. Instruction also comprises reviewing and commenting on each other's' drafts in pairs with the focus on the conceptual relation and the audience and revising one's own draft individually using the received feedback.

5.6.3 Connecting writing-to-learn to language policy in schools

Klein and Boscolo (2016) mention that writing is a way of learning that is not frequently used in education. For introducing writing-to-learn in classroom practice, a connection to related initiatives can be sought. This connection can be found in the schools' language policy, which is a school team's structural and strategic approach to adjust classroom practice to students' language learning needs, for the benefit of their overall development and improvement of their educational outcomes (Hajer & Meestringa, 2015). Language policy concerns academic and disciplinary language to enable students to understand and discuss subject matter.

Writing-to-learn can be introduced in schools with a well-functioning language policy that is aimed at familiarizing students with academic language in various disciplines. This may be a good starting point for introducing writing-to-learn. After all, academic language skills are a prerequisite for learning by writing. Writing-to-learn tasks require students to direct their texts to a lay audience and therefore translate academic contents into everyday language which can be understood by their audience. In this way, writing-to-learn tasks not only challenge students to acquire new insights into subject matter, but also to test and deepen their understanding of

academic language. Performing writing-to-learn tasks requires students to apply their academic language skills focused on the context of the school's language policy. Therefore, writing-to-learn can add to and reinforce the effects of language policy in schools.

Although in the previous sections it appears that implementing writing-to-learn in regular education is difficult and (in the beginning) quite time consuming, this extra effort for students' learning is certainly worthwhile, as has been demonstrated by the studies in this dissertation. The studies show that writing-to-learn can be used in various types of education, in higher education as well as in prevocational education. They also show that writing-to-learn can be used in divergent disciplines, such as biology and mathematics. Therefore, we can assume it can be introduced in various other disciplines as well.

References

Sources marked with (*) were selected for the review (Chapter 2)

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

Akkus, R., & Hand, B. (2011). Examining teachers' struggles as they attempt to implement dialogical interaction as part of promoting mathematical reasoning within their classrooms. *International Journal of Science and Mathematics Education 9*, 975–998.

Akkus, R., Gunel, M., & Hand, B. (2007). Comparing an inquiry based approach known as the Science Writing Heuristic to traditional science teaching practices: Are there differences? *International Journal of Science Education*,*29*(14), 1745-1765. https://doi.org/10.1080/09500690601075629

Applebee, A. N. (1984). Writing and reasoning. *Review of Educational Research, 54,* 577–596. https://doi.org/10.3102/00346543054004577

*Atasoy, Ş., & Küçük, O. (2020). Development of eighth grade students' epistemological beliefs through writing-to-learn activities. *Journal of Science Learning 3*(2), 57-66.

Baaijen, V.M., (2012), *The development of understanding through writing* (Doctoral dissertation.). Groningen Dissertations in Linguistics (107).

Baaijen, V.M., & Galbraith, D. (2018). Discovery through writing: Relationships with writing processes and text quality. *Cognition and Instruction*, *36*(3), 199-223. https://doi.org/10.1080/07370008.2018.1456431

Baker, W.P., Barstack, R., Clark, D., Hull, E., Goodman, B., Kook, J., Kraft, K., Ramakrishna, P., Roberts, E., Shaw, J., Weaver, & Michael Lang, M. (2008). Writing-to-learn in the inquiry classroom: Effective strategies from middle school science and writing teachers. *The Clearing House*, *81,* 105–108. https://doi.org/10.3200/TCHS.81.3.105-108

*Balgopal, M. M., Wallace, A. M., & Dahlberg, S. (2012). Writing to learn ecology: A study of three populations of college students. *Environmental Education Research*, *18*, 67–90. https://doi.org/10.1080/13504622.2011.576316

Bangert-Drowns, R. L., Hurley, M. M., & Wilkinson, B. (2004). The effects of school-based writing-to-learn interventions on academic achievement: a meta-analysis. *Review of Educational Research, 74*, 29–58. https://doi.org/10.3102/00346543074001029

Bazerman, C. (2005). A response to Anthony Fleury's "Liberal education and communication against the disciplines": A view from the world of writing. *Communication education*, *54*(1), 86-91. https://doi.org/10.1080/03634520500077099

Beal, C.R. (1990). The development of text evaluation and revision skills. *Child Development*, *61*, 247-258.

Bean, J. C. (2011). *Engaging ideas: The professor's guide to integrating writing, thinking and active learning in the classroom.* Jossey-Bass.

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Erlbaum.

Berthold, K., Nückles, M., & Renkl, A. (2007). Do learning protocols support learning strategies and outcomes? The role of cognitive and metacognitive prompts. *Learning and Instruction, 17*, 564–577. https://doi.org/10.1016/j.learninstruc.2007.09.007

Bos, D., Bouw, J., Brinkman, H. De Bruijn, I., Doekes, W., Van Dijk, B, Franken, E., Van de Giessen, C., Koning, A., Van der Linden, T., Roorda, G., Sinkeldam, R., Schaberg, G., & Stolwijk, R. (2007). *Moderne Wiskunde*: vwo-bovenbouw deel B1, B2, B3, D1, D3. 9e editie. Noordhoff.

*Boscolo, P., & Carotti, L. (2003). Does writing contribute to improving high school students' approach to literature? *L1, Educational Studies in Language and Literature, 3*, 197–224. Kluwer Academic Publishers. https://doi.org/10.1023/B:ESLL.0000003685.30925.c4 -

Boyle, G. J. (1991). Does item homogeneity indicate internal consistency or item redundancy in psychometric scales? *Personality and individual differences 12* (3), 211-328. *https://doi.org/*10.1016/0191-8869(91)90115-R

Boyles, M. P., Killian, P. W., & Rileigh, K. K. (1994). Learning by writing in Introductory psychology. *Psychological Reports, 75,* 563–568. https://doi.org/10.2466/pr0.1994.75.1.563

Britton, J. (1982). Shaping at the point of utterance. In G.M. Pradl (ed.) *Prospect and retrospect*: *Selected essays of James Britton*, 139-145. Boynton/Cook.

Brozo, W.G., & Crain, S. (2018). Writing in math: A disciplinary literacy approach. *The Clearing House, 91*(1*),* 7–13. https://doi.org/10.1080/00098655.2017.1342435

Campbell, N.A. et al. (2011). *Biology.* Pearson.

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences.* Academic press.

*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

Countryman, J. (1992). *Writing to learn mathematics: Strategies that work*. Heinemann.

Crocker, D.A. (1992). Educational reflections. *The Amatic Review, 13*(2*),* 66–68. https://doi.org/10.1177/174114329202000201

Darwin, C. (2010). *Het ontstaan van soorten door natuurlijke selectie ofwel het bewaard blijven van rassen die in het voordeel zijn in de strijd om het bestaan.* Amstel Uitgevers (Olympus).

Dunn, Loyd M., & Dunn, Leota M. Dunn (2005*). Peabody picture vocabulary test – III-NL*. L. Schlichting (Dutch version). Harcourt Test Publishers.

Diaz Eaton, C., & Wade, S. (2014). Collaborative learning through formative peer review with technology. *Primus, 24*(6), 529–543. https://doi.org/10.1080/10511970.2014.881442

Emig, J. (1977). Writing as a mode of learning. *College Composition and Communication, 28,* 122–128. https://doi.org/10.2307/356095

Ericsson, K.A., & Simon, H.A. (1993). *Protocol analysis*. MIT Press.

Faber, J. E., Morris, J.D., & Liebermann, M.G. (2000). The effect of note taking on ninth grade students' comprehension. *Reading Psychology*, *21,* 257–270.

*Finkenstaedt-Quinn, S. A., Halim, A. S., Chambers, T. G., Moon, A., Goldman, R. S., Ruggles Gere, A., & Shultz, G. V. (2017). Investigation of the influence of a writing-to-learn assignment on student understanding of polymer properties. *Journal of Chemical Education, 94,* 1610–1617. https://doi.org/10.1021/acs.jchemed.7b00363

Galbraith, D. (1992). Conditions for discovery through writing. *Instructional Science, 21,* 45–72. https://doi.org/10.1007/BF00119655

Galbraith, D. (1999). Writing as a knowledge-constituting process. In M. Torrance, & D. Galbraith (Eds.), *Knowing what to write: Conceptual processes in text production,* 139–160. Amsterdam University Press.

Galbraith, D. (2009). Writing as discovery. *British Journal of Educational Psychology Monograph, Series II-6,* 5–26. https://doi.org/10.1348/978185409X421129

Galbraith, D. (2015). Conditions for writing to learn. *Journal for Writing Research, 7*(1), 215–266. https://doi.org/10.17239/jowr-2015.07.01.09

Galbraith, D., & Torrance, M. (2004). Revision in the context of different drafting strategies. In L. Allal, L. Chanquoy, & P. Largy (Eds.), *Revision: Cognitive and instructional processes, 63*–86. Kluwer Academic Publishers. https://doi.org/10.1007/978-94-007-1048-15

Graham, S., Kiuhara, S. A., & McKay, 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

*Granado-Peinado, M., Mateos, M., Martin, E., & Cuevas, I. (2019). Teaching to write collaborative argumentative syntheses in higher education. *Reading and Writing 32,* 2037-2058. https://doi.org/10.1007/s11145-019-09939-6

*Gunel, M., Hand, B., & Gunduz, S. (2006). Comparing student understanding of quantum physics when embedding multimodal representations into two different writing formats: Presentation format versus summary report format. *Wiley InterScience,* 1092-1112. https://doi.org/10.1002/sce.20160

*Gunel, M., Hand, B., & McDermott, M. A. (2009). Writing for different audiences: Effects on high-school students' conceptual understanding of biology. *Learning and Instruction 19,* 354–376. https://doi.org/10.1016/j.learninstruc.2008.07.001

Gunel, M., & Yesildag-Hasancebi, F. (2016). Modal representations and their role in the learning process: A theoretical and pragmatic analysis. *Educational Sciences: Theory and Practice*, *16 (1),* 109–126.

Hajer, M., & Meestringa, T. (2015). *Handboek Taalgericht Vakonderwijs*. Coutinho.

Halliday, M.A.K., & Martin, J.R. (1993). *Writing science: Literacy and discursive power*. Routledge.

Hand, B., Chen, Y., & Suh, J. K. (2021). Does a knowledge generation approach to learning benefit students?: A systematic review of research on the Science Writing Heuristic approach. *Educational Psychology Review 33*, 535-577. https://doi.org/10.1007/s10648-020-09550-0

*Hand, B., Gunel, M., & Ulu, C. (2009). Sequencing embedded multimodal representations in a writing to learn approach to the teaching of electricity. *Journal of Research in Science Teaching, 46*(3)*,* 225–247. https://doi.org/10.1002/tea.20282

*Hand, B., Hohenshell, L., & Prain, V. (2004). Exploring students' responses to conceptual questions when engaged with planned writing experiences: A study with year 10 science students. *Journal of Research in Science Teaching, 41*(2), 186–210. https://doi.org/10.1002/tea.10128

*Hand, B., Hohenshell, L., & Prain, V. (2007). Examining the effect of multiple writing tasks on year 10 biology students' understanding of cell and molecular biology concepts. *Instructional Science, 35*, 343–373. https://doi.org/10.1007/s11251-006-9012-3

*Hand, B., Wallace, C.W., & Yang, E. (2004). Using a Science Writing Heuristic to enhance learning outcomes from laboratory activities in seventh-grade science: Quantitative and qualitative aspects. *International Journal of Science Education, 26*(2), 131–149. https://doi.org/10.1080/0950069032000070252

*Hand, B., Yang, O. E., & Bruxvoort, C. (2007). Using writing-to-learn science strategies to improve year 11 students' understanding of stoichiometry. *International Journal of Science and Mathematics Education, 5,* 125–143. https://doi.org/10.1007/s10763-006-9029-8

Hayes, J. R., & Flower, L. S. (1980). Identifying the organization of writing processes. In L. W. Gregg, & E. R. Steinberg (Eds.). *Cognitive processes in writing,* 3–27. Erlbaum.

*Hohenshell, L. M., & Hand, B. (2006). Writing-to-learn strategies in secondary school cell biology: A mixed method study. *International Journal of Science Education*, *28*(2-3), 261–289. https://doi.org/10.1080/09500690500336965

Hübner, S., Nückles, M., & Renkl, A. (2010). Writing learning journals: Instructional support to overcome learning-strategy deficits. *Learning and Instruction*, *20*, 18-29. https://doi.org/10.1016/j.learninstruc.2008.12.001

*Hunter, K., & Tse, H. (2013). Making disciplinary writing and thinking practices an integral part of academic content teaching. *Active Learning in Higher Education, 14* (3), 227–239. https://doi.org/10.1177/1469787413498037

*Kabataş Memiş, E., & Öz, M. (2017). The impact of using representation modes within writing to learn activities on the scientific process skills of the fifth-grade students. *Journal of Education and Training Studies, 5(2)*, 31– 42. https://doi.org/10.11114/jets.v5i2.2001

Kasparek, R. F. (1993). *Effects of integrated writing on attitude and algebra performance of high school students.* Doctoral dissertation. *Abstracts International*, *54*(08), 2931A. (UMI No. 9402483).

Kenney, R, Shoffner, M., & Norris, D. (2014). Reflecting on the use of writing to promote mathematical learning: An examination of preservice mathematics teachers' perspectives. *The Teacher Educator, 49*(1), 28–43. https://doi.org/10.1080/08878730.2013.848002

*Kieft, M., Rijlaarsdam, G., & Van den Bergh, H. (2006). Writing as a learning tool: Testing the role of students' writing strategies. *European Journal of Psychology of Education, XXI* (1), 17–34.

*Kieft, M., Rijlaarsdam, G., & Van den Bergh, H. (2008). An aptitude treatment interaction approach to writing-to-learn. *Learning and Instruction,18*, 379–390. https://doi.org/10.1016/j.learninstruc.2007.07.004

Kieft, M., Rijlaarsdam, G., Galbraith, D., & Van den Bergh, H.(2007). The effects of adapting a writing course to students' writing strategies. *British Journal of Educational Psychology 77*, 565-578. https://doi.org/10.1348/096317906X120231

Kiuhara, S.A., Graham, S., & Hawken, L.S. (2009). Teaching Writing to High School Students: A National Survey. *Journal of Educational Psychology*, *101* (1), 136–160. https://doi.org/ 10.1037/a0013097

Klein, P.D. (1999). Reopening inquiry into cognitive processes in writing-to-learn. *Educational Psychology Review, 11*(3), 203–270. https://doi.org/10.1023/A:1021913217147

Klein, P. D. (2004). Constructing scientific explanations through writing. *Instructional Science, 32,* 191-231.

Klein, P. D., & Boscolo, P. (2016). Trends in research on writing as a learning activity. *Journal of Writing Research, 7*(3), 311-350. https://doi.org/10.17239/jowr-2016.07.03.01

*Klein, P. D., & Ehrhardt, J. S. (2015). The effects of discussion and persuasion writing goals on reasoning, cognitive load, and learning. *Alberta Journal for Educational Research, 61*(1), 40–64.

*Klein, P. D., & Kirkpatrick, L. C. (2010). A framework for content area writing: Mediators and moderators. *Journal of Writing Research, 2* (1), 1–46. https://doi.org/10.17239/jowr-2010.02.01.1

*Klein, P. D., & Rose, M. A. (2010). Teaching argument and explanation to prepare junior students for writing to learn. *Reading Research Quarterly, 45*(4), 433–461. https://doi.org/10.1598/RRQ.45.4.4

*Klein, P. D., & Samuels, B. (2010). Learning about plate tectonics through argument writing. *The Alberta Journal of Educational Research, 56* (2), 196–217

Klein, P.D., Boscolo, P., Gelati, C., & Kirkpatrick, L.C. (2014). New directions in writing as a learning activity. In G. Rijlaarsdam (Series Ed.), & P.D. Klein, P. Boscolo, L.C. Kirkpatrick, C. Gelati (Vol. Eds.), *Studies in Writing: Vol. 28, Writing as a learning activity,* 1-14. Brill.

Langer, J.A., & Applebee, A.N. (1987). *How writing shapes thinking: A study of teaching and learning.* NCTE.

Linnemann, M., & Stephany, S. (2014). Supportive writing assignments for less skilled writers in the mathematics classroom. In G. Rijlaarsdam (Series Ed.), & P. D. Klein, P. Boscolo, L. C. Kirkpatrick, C. Gelati (Vol. Eds.), *Studies in Writing: Vol. 28, Writing as a learning activity,* 66–93. Brill. https://doi.org/10.1163/9789004265011_005

*Martinez, I., Mateos, M., Martin, E., & Rijlaarsdam, G. (2015). Learning history by composing synthesis texts: Effects of an instructional program 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

Mason, L., (1998). Sharing cognition to construct scientific knowledge in school context: The role of oral and written discourse. *Instructional Science, 26,* 359-389.

* Mateos, M., Martin, E., Cuevas, I., Villalón, R., Martinez, I., & González-Lamas, J. (2018). Improving written argumentative synthesis by teaching the integration of conflicting information from multiple sources. *Cognition and Instruction 36*(2), 119-138.https://doi.org/10.1080/07370008.2018.1425300

*McDermott, M.A., & Hand, B. (2013). The impact of embedding multiple modes of representation within writing tasks on high school students' chemistry understanding. *Instructional Science, 41,* 217–246. https://doi.org/10.1007/s11251-012-9225-6

Miller, D. M., Scott, C. E., & McTigue, M. (2018). Writing in the secondary level disciplines: A systematic review of context, cognition and content. *Educational Psychology Review, 30,* 83–120. https://doi.org/10.1007/s10648-016-9393-z

*Nam, J., Choi A., & Hand, B. (2011). Implementation of the science writing heuristic (SWH) approach in 8th grade science classrooms. *International Journal of Science and Mathematics Education, 9*, 1111–1133. https://doi.org/10.1007/s10763-010-9250-3

Nevid, J.S., Ambrose, M.A., & Pyun, Y.S. (2017). Effects of higher and lower level writing to learn assignments on higher and lower level examination questions. *Teaching of Psychology, 44*(4), 324-329. https://doi.org/10.1177/0098628317727645

*Nevid, J.S., Pastva, A., & McClelland, N. (2012). Writing-to-learn assignments in introductory psychology: Is there a learning benefit? *Teaching of Psychology, 39*(4), 272–275. https://doi.org/10.1177/0098628312456622

Newell, G. E. (1984). Learning from writing in two content areas: A case study/protocol analysis. *Research in the Teaching of English, 18(3)*, 265–287.

*Nückles, M., Hübner, S., Dümer, S., & Renkl, A. (2010). Expertise reversal effects in writing-to-learn. *Instructional Science, 38*, 237–258. https://doi.org/10.1007/s11251-009-9106-9

Nückles, M., Hübner, S. & Renkl, A. (2009). Enhancing self-regulated learning by writing learning protocols. *Learning and Instruction, 19,* 259-271.

Pander Maat, H. (1994). *Tekstanalyse: Een pragmatische benadering.* Martinus Nijhoff.

Page, M.J., et al. (2020). The Prisma 2020 statement: an updated guideline for reporting systematic reviews. https://doi.org/10.1136/bmj.n71

Prain, V. (2006). Learning from writing in secondary science: Some theoretical and practical implications. *International Journal of Science Education. 28(2–3)*, 179–201. https://doi.org/10.1080/09500690500336643

Reichard, L. A. (2011a). *Getal & ruimte: wi 3 vmbo-K deel 2.* EPN

Reichard, L.A. (2011b). *Getal & ruimte: wi 4 vmbo-K deel 1.* EPN.

*Ritchie, S.M., Tomas, L., & Tones, M. (2011). Writing stories to enhance scientific literacy. *International Journal of Science Education*, *33*(5), 685–707. https://doi.org/10.1080/09500691003728039

Rivard, L.P. (1994). A Review of writing to learn in science: Implications for practice and research. *Journal of Research in Science Teaching, 31*(9), 969–983. https://doi.org/10.1002/tea.3660310910

*Rivard, L.P. (2004). Are language-based activities in science effective for all students, including low achievers? *Science Education, 88*, 420–442. https://doi.org/10.1002/sce.10114

*Rivard, L.P., & Straw, S.B. (2000). The effect of talk and writing on learning science: An exploratory study. *Science Education, 84,* 566–593.

Rose, D. (2008). Writing as a linguistic mastery: The development of genre-based literacy pedagogy. In D. Myhill, D. Beard, M. Nystrand, & J. Riley (eds.) *Handbook of Writing Development*, 151- 170.Sage. https://doi.org/10.4135/9780857021069

*Sampson, V., & Phelps Walker, J. (2012). Argument-Driven inquiry as a way to help undergraduate students write to learn by learning to write in chemistry. *International Journal of Science Education, 34(*10), 1443–1485. https://doi.org/10.1080/09500693.2012.667581

Schleppegrell, M.J. (2004). *The language of schooling. A functional linguistics perspective.* Erlbaum. https://doi.org/10.4324/9781410610317

Skemp, R. R. (1987). *The psychology of learning mathematics.* Routledge.

Smits, G. (2008). *Biologie voor jou 1 Vmbo-kgt*. Malmberg.

*Stewart, T.L., Myers, A.C., & Culley, M.R. (2010). Enhanced learning and retention through "writing to learn" in the psychology classroom. *Teaching of Psychology, 37*, 46–49. https://doi.org/10.1080/00986280903425813

Sijtsma, K. (2009). On the use, the misuse and the very limited usefulness of Cronbach's alpha. *Psychometrika, 74* (1), 107-120. https://doi.org/10.1007/S11336-008-9101-0

Taber, K.S. (2018). The use of Cronbach's Alpha when developing and reporting research instruments in science education. *Research in Science Education, 48*, 1273-1296. https://doi.org/10.1007/s11165-016-9602-2

Thijssen, J., Van der Schoot, F., & Hemker, B. (2011). *Balans van het biologie-onderwijs aan het einde van de basisschool 4.* Cito.

Van Berchum, C. (2010). *Netwerk: 3B vmbo-kader*. Noordhoff.

Van Dijk, A., Van Gelderen, A., & Kuiken, F., (2022a). Which types of instruction in wrting-to-learn lead to insight and topic knowledge in different disciplines? A review of empirical studies. *Review of Education*. https://doi.org/10.1002/rev3.3359

Van Dijk, A., Van Gelderen, A., & Kuiken, F., (2022b). Effects of instruction in writing-to-learn on low achieving adolescents in biology and mathematics classes. *Education Sciences 12*, 595. https://doi.org/10.3390/edusci12090595

Van Dijk, A., Van Gelderen, A., & Kuiken, F., (2023). Writing-to-learn in biology and mathematics teacher education. Promoting students' topic knowledge and insight. *Frontiers in Education, 8*. |https://doi.org/10.3389/feduc.2023.1094156

*Van Drie, J., Braaksma, M., & Van Boxtel, C. (2015). Writing in history: Effects of writing instruction on historical reasoning and text quality. *Journal of Writing Research*, *7(*1), 123–156. https://doi.org/10.17239/jowr- 2015.07.01.06

Van Gelderen, A. (1997). Elementary students' skills in revising: Integrating quantitative and qualitative analysis. *Written Communication, 14*(3), 360-397. https//doi.org/ 10.1177/0741088397014003003

Van Gelderen, A., & Blok, H. (1991). De praktijk van het stelonderwijs in de groepen 7 en 8 van de basisschool; observaties en interviews. *Pedagogische Studien, 68 (*4), 159-175.

Van Someren, M.W., Barnard, Y. F., & Sandberg, J.A.C. (1994). *The think aloud method: A practical guide to modelling cognitive processes.* Academic Press Limited.

Veel, R. (1999). Language, knowledge and authority in school mathematics. In F. Christie (Ed.), *Pedagogy and the shaping of consciousness*: *Linguistic and social processes.* Cassell.

Wallace, C.S., Hand, B., & Prain, V. (2007). *Writing and learning in the science classroom.* Springer.

*Wäschle, K, Gebhardt, A, Oberbusch, E.M., & Nückles, M. (2015). Journal writing in science: Effects on comprehension, interest, and critical reflection. *Journal of Writing Research, 7(1),* 41–64. https://doi.org/10.17239/jowr-2015.07.01.03

*Winstead Fry, S., & Villagomez, A. (2012). Writing to learn benefits and limitations. *College Teaching, 60*, 170–175. https://doi.org/10.1080/87567555.2012.697081

*Wissinger, D.R., & De La Paz, S. (2016). Effects of critical discussions on middle School students' written historical arguments. *Journal of Educational Psychology, 108*(1)*,* 43–59. https://doi.org/10.1037/edu0000043

*Wong, B.Y.L., Kuperis, S., Jamieson, D., Keller, L., & Cull-Hewitt, R. (2002). Effects of guided journal writing on students' story understanding. *The Journal of Educational Research, 95*(3), 179–191. https://doi.org/10.1080/00220670209596588

*Yildiz, E. & Akdag, S. (2021). The effects of cooperative learning and writing to learn applications on academic achievement. *International Journal of Progressive Education 17*(1), 196-209. https://doi.org/10.29329/ijpe.2020.329.13

Zeiss, F. (1999). *Natuurlijke historiën*: *Geschiedenis van de biologie van Aristoteles tot Darwin.* Boom.

Appendices

Appendix A The process of writing-to-learn: all coded frequencies for biology teacher education (continued on the next page)

Proportioned means and standard deviations of frequencies of the codes in experimental condition (N=6) and control group (N=6), and effect sizes.

		Experimental		Control		Effect size
Codes		Mean	(SD)	Mean	(SD)	Cohen's *d*
PLANNING: orienting						
1 Read. assignment		.00	(.00)	.00	(.00)	.00
2 Thinking about task approach		.05	(.04)	.05	(.02)	.00
PLANNING: generating						
3 Generating content		.00	(.00)	.00	(.00)	.00
4 Using knowledge about audience	**I**	.00	(.00)	.00	(.00)	.00
5 Using source text for new content		.18	(.08)	.19	(.14)	- .10
PLANNING: selecting						
6 Thinking about content selection	**I**	.03	(.04)	.04	(.04)	- .27
7 Selecting content		04	(.03)	.04	(.03)	.00
PLANNING: organizing						
8 Ordering ideas		.00	(.00)	.00	(.00)	.00
9 Thinking about text structure		.00	(.01)	.00	(.00)	.00
FORMULATING						
10 Thinking about formulating	**I**	.02	(.02)	.02	(.03)	.00
11 Formulating		.25	(.05)	.30	(.18)	- .41
12 Revising while formulating	**I**	.03	(.03)	.02	(.01)	.49
13 Repeating formulation		.02	(.01)	.01	(.01)	1.10
14 Revising after finishing an utterance	**I**	.03	(.04)	.01	(.01)	.75
MONITORING						
15 Rereading own text	**I**	.08	(.09)	.03	(.03)	.2
16 Rereading assignment		.00	(.00)	.00	(.00)	.00
17 Using source text for correct understanding		.01	(.01)	.04	(.06)	- .76
18 Rethinking task approach	**I**	.01	(.01)	.01	(.01)	.00
19 Checking task completion		.01	(.01)	.01	(.01)	.00
EVALUATION						
20 Commenting own phrases		.03	(.03)	.01	(.02)	.86
21 Commenting text structure		.00	(.00)	.00	(.01)	.00

	Experimental		Control		Effect size
Codes	Mean	(SD)	Mean	(SD)	Cohen's *d*
22 Commenting task performance incl. writing process	.03	(.02)	.04	(.03)	- .37
23 Commenting on assignment	.00	(.00)	.00	(.00)	.00
24 Commenting on source text	.04	(.03)	.04	(.03)	.00
OTHER ACTIVITIES					
25 Pause for thinking	.05	(.05)	.04	(.04)	.24
26 Long pause (more than 50 seconds)	.01	(.02)	.00	(.00)	.77
27 Transcribing	.02	(.02)	.02	(.02)	.00
28 Not task related remarks	.14	(.07)	.15	(.12)	- .11
29 Expression of uncertainty	.00	(.00)	.01	(.01)	- 1.55

I*=Indicator for writing-to-learn*

Appendix B The process of writing-to-learn: all coded frequencies for mathematics teacher education (continued on the next page)

Proportioned means and standard deviations of frequencies of codes in experimental condition (N=8) and control group (N=7), and effect sizes

	Experimental		**Control**		**Effect size**
Codes	Mean	(SD)	Mean	(SD)	Cohen's d
PLANNING: orienting					
1 Read. assignment	01	(.01)	.02	(.04)	- .32
2 Thinking about task approach	.03	(.05)	.04	(.04)	- .24
PLANNING: generating					
3 Generating content	.00	(.01)	00	(.00)	.00
4 Using knowledge about audience	.04	(.03)	.02	(.02)	.83
5 Using source text for new content	.00	(.00)	.00	(.00)	.00
PLANNING: selecting					
6 Thinking about content selection	.03	(.05)	.01	(.01)	.55
7 Selecting content	.04	(.04)	.01	(.01)	.86
PLANNING: organizing					
8 Ordering ideas	.01	(.02)	.00	(.00)	.66
9 Thinking about text structure	.01	(.01)	.01	(.01)	.00
FORMULATING					
10 Thinking about formulating	.03	(.02)	.04	(.07)	- .25
11 Formulating	.39	(.09)	.43	(.15)	- .31
12 Revising while formulating	.03	(.04)	.01	(.02)	.38
13 Repeating formulation	.02	(.02)	.02	(.02)	.00
14 Revising after finishing an utterance	.02	(.02)	.02	(.02)	.00
MONITORING					
15 Rereading own text	.12	(.09)	.04	(.06)	1.27
16 Rereading assignment	.00	(.00)	.01	(.01)	- .26
17 Using source text for correct understanding	.00	(.00)	.00	(.00)	.00
18 Rethinking task approach	.01	(.01)	.01	(.02)	.00
19 Checking task completion	.02	(.03)	.00	(.00)	.83
EVALUATION					
20 Commenting own phrases	.04	(.04)	.02	(.03)	.73
21 Commenting text structure	.00	(.00)	.00	(.00)	.00
22 Commenting task performance incl. writing process	.03	(.03)	.07	(.10)	- 1.43
23 Commenting on assignment	.00	(.00)	.00	(.00)	.00
24 Commenting on source text	.00	(.00)	.00	(.00)	.00
OTHER ACTIVITIES					
25 Pause for thinking	.11	(.06)	.10	(.10)	.13
26 Long pause (more than 50 seconds)	.01	(.01)	.01	(.01)	.00

	Experimental		**Control**		**Effect size**
Codes	Mean	(SD)	Mean	(SD)	Cohen's d
27 Transcribing	.02	(.03)	.01	(.02)	.47
28 Not task related remarks	.12	(.02)	.21	(.15)	- 2 .47
29 Expression of uncertainty	.01	(.05)	.00	(.00)	.10

I*=Indicator for writing-to-learn*

Appendix C The process of writing-to-learn: all coded frequencies for grade 7 biology

Proportioned means and standard deviations of frequencies of codes in experimental (N=5) and control group (N=6), and effect sizes.

		Experimental		Control		Effect size
Codes		Mean	(SD)	Mean	(SD)	Cohen's d
PLANNING: orienting						
1 Read. assignment		.03	(.04)	.05	(.06)	- .39
2 Thinking about task approach		.03	(.04)	.00	(.01)	1.03
PLANNING: generating						
3 Generating content		.03	(.06)	.04	(.07)	- .15
4 Using knowledge about audience	**I**	.00	(.00)	.00	(.01)	.00
5 Using source text for new content		.00	(.00)	.00	(.00)	.00
PLANNING: selecting						
6 Thinking about content selection	**I**	.09	(.07)	.02	(.03)	1.30
7 Selecting content		.01	(.02)	.00	(.01)	.63
PLANNING: organizing						
8 Ordering ideas		.00	(.00)	.00	(.00)	.00
9 Thinking about text structure		.01	(.01)	.00	(.00)	1.41
FORMULATING						
10 Thinking about formulating	**I**	.01	(.02)	.01	(.02)	.00
11 Formulating		.21	(.20)	.28	(.13)	- .42
12 Revising while formulating	**I**	.00	(.01)	.02	(.03)	- .89
13 Repeating formulation		.00	(.00)	.00	(.00)	.00
14 Revising after finishing an utterance	**I**	.01	(.01)	.01	(.01)	.00
MONITORING						
15 Rereading own text	**I**	.02	(.01)	.02	(.03)	.00
16 Rereading assignment		.02(	.03)	.03	(.03)	- .33
17 Using source text for correct understanding		.00	(.00)	.00	(.00)	.00
18 Rethinking task approach	**I**	.01	(.01)	.01	(.02)	.00
19 Checking task completion		.01	(.02)	.03	(.01)	-1.26
EVALUATION						
20 Commenting own phrases		.00	(.00)	.00	(.00)	.00
21 Commenting text structure		.00	(.00)	.00	(.00)	.00
22 Commenting task performance incl. writing process		.03	(.03)	.03	(.03)	.00

	Experimental		Control		Effect size
Codes	Mean	(SD)	Mean	(SD)	Cohen's d
23 Commenting on assignment	.00	(.00)	.00	(.00)	.00
24 Commenting on source text	.00	(.00)	.00	(.00)	.00
OTHER ACTIVITIES					
25 Pause for thinking	.12	(.10)	.16	(.14)	.24
26 Long pause (more than 50 seconds)	.01	(.01)	.00	(.00)	.77
27 Transcribing	.12	(.03)	.07	(.02)	.00
28 Not task related remarks	.23	(.15)	.22	(.09)	- .11
29 Expression of uncertainty	.00	(.00)	.00	(.00)	- 1.55

I*=Indicator for writng-to-learn*

Appendix D The process of writing-to-learn: all coded frequencies for grade 10 mathematics (continued on the next page)

Proportioned means and standard deviations of frequencies of codes in experimental (N=4)) and control group (N=4), and effect sizes

	Experimental		Control		Effect size
Codes	Mean	(SD)	Mean	(SD)	Cohen's d
PLANNING: orienting					
1 Read. Assignment	.03	(.03)	.04	(.03)	-.33
2 Thinking about task approach	.02	(.02)	.02	(.02)	.00
PLANNING: generating					
3 Generating content	.00	(.00)	.00	(.00)	.00
4 Using knowledge about audience **I**	.03	(.05)	.01	(.02)	.53
5 Using source text for new content	.01	(.02)	.00	(.00)	.71
PLANNING: selecting					
6 Thinking about content selection **I**	.00	(.00)	.01	(.05)	-.29
7 Selecting content	.00	(.00)	.03	(.05)	-.85
PLANNING: organizing					
8 Ordering ideas	.00	(.00)	.00	(.01)	.00
9 Thinking about text structure	.00	(.00)	.00	(.00)	.00
FORMULATING					
10 Thinking about formulating **I**	.00	(.00)	.00	(.00)	.00
11 Formulating	.35	(.21)	.28	(.05)	.46
12 Revising while formulating **I**	.00	(.00)	.00	(.00)	.00
13 Repeating formulation	.00	(.01)	.00	(.01)	.00
14 Revising after finishing an utterance **I**	.02	(.02)	.00	(.00)	1.41
MONITORING					
15 Rereading own text **I**	.07	(.13)	.03	(.03)	.42
16 Rereading assignment	.01	(.02)	.00	(.00)	.71
17 Using source text for correct understanding	.00	(.00)	.00	(.00)	.00
18 Rethinking task approach **I**	.16	(.19)	.10	(.09)	.40
19 Checking task completion	.01	(.02)	.02	(.02)	- .50
EVALUATION					
20 Commenting own phrases	.00	(.00)	.00	(.00)	.00
21 Commenting text structure	.00	(.00)	.00	(.00)	.00
22 Commenting task performance incl. writing process	.03	(.03)	.01	(.02)	.78
23 Commenting on assignment	.01	(.01)	.00	(.00)	1.41
24 Commenting on source text	.00	(.00)	.00	(.00)	.00
OTHER ACTIVITIES					
25 Pause for thinking	.08	(.03)	.13	(.09)	- .76
26 Long pause (more than 50 seconds)	.00	(.00)	.00	(.00)	.00
27 Transcribing	.03	(.03)	.06	(.04)	- .85
28 Not task related remarks	.13	(.10)	.25	(.08)	- .33

Codes	Experimental Mean	(SD)	Control Mean	(SD)	Effect size Cohen's d
29 Expression of uncertainty	.01	(.02)	.00	(.00)	.71

I*=Indicator for writng-to-learn*

Appendix E. Model text for grade 7 low achieving biology students.

Genre: explanation. Conceptual relation: comparison.

A correct posture

For people who often must lift heavy things because of their profession, a correct posture is very important. To explain this, we compare two moving men to each other. Imagine, that both carry a very heavy box, full of books.
Moving man 1 holds the box against his stomach and carries it with bent arms. Moving man 2 lifts the box **much higher**, above the shoulders, and carries it with stretched arms. For Moving man 2 the job is **much heavier**, because he stretches his arms and cannot lean the box against his belly. You have probably noticed that your arm muscles must work **harder** when you lift a box with stretched arms **than** when you lift a box with bent arms.
Mover 1 keeps his back straight when lifting. That's easier on his back muscles. Mover 2 does not keep his back straight, but obliquely backwards. His back muscles have to pull harder to keep his back up than the back muscles of mover 1. Mover 1 has the correct posture

Note: bold is used to support students at recognizing the comparisons. In the last paragraph students must find the comparison without support.

Summary

Effects of instruction in writing-to-learn in different disciplines and types of education

Students in all types of education often have difficulty acquiring topic knowledge and insight into subject matter of various disciplines. They do not understand or cannot remember theory from their textbook (Sampson & Phelps Walker, 2012; Hunter & Tse, 2013; Finkenstaedt-Quinn et al., 2017). Usually, students learn theory by studying textbooks, by listening closely to their teacher's explanations, or by discussing subject matter with peers. However, this is often not enough to get a grip on subject matter. Writing-to-learn can help; however, this way of learning is not yet much applied in education. In writing-to-learn, the aim is to promote learning by writing texts about subject matter. This allows students to reread their written knowledge and to reflect whether they have displayed the disciplinary content correctly. Writing-to-learn is writing with the aim to gain insight into subject matter and to acquire topic knowledge.

Writing-to-learn has been studied since the 1980s. At first, it was thought that writing in itself brings about the learning process (strong text theory), but this appeared not always to be the case (Klein, 1999). A necessary condition to stimulate students' thinking is instruction of cognitive and metacognitive strategies in writing tasks, leading to new insights into and knowledge of subject matter (Bangert-Drowns et al., 2004). An example of instruction of a cognitive strategy in a writing task is: 'first try to organize the content of the text you are going to write in a meaningful way'. An example of instruction of a metacognitive strategy is: ‘write down which topics from the lesson you do not understand yet’ (Nückles et al., 2009). The instructions incite students to reflect. Reflection can support them to gain new insights into and topic knowledge of subject matter (Bereiter & Scardamalia, 1987).

The dissertation starts with a literature study by means of which was investigated which types of instruction in writing tasks have positive effects on learning in different types of education and disciplines. Subsequently, four quasi-experimental studies were carried out into the effects of a certain type of instruction in different types of education and disciplines. Following each of the four quasi-experiments, a think aloud study of the writing-to-learn process was conducted. A total of nine studies have been conducted, which are discussed below.

Literature review

The aim of the literature review in this thesis (Chapter 2) is to provide an overview of research on effects of different types of instruction in writing-to-learn. Bangert-Drowns' et al. (2004) suggestion to investigate whether and how instruction of (meta)cognitive strategies in writing tasks can lead to new insights and new topic knowledge has been taken to heart in a large number of studies. Some review studies report positive effects of writing-to-learn instruction on learning processes (Miller et al., 2018; Graham et al., 2020; Hand et al., 2021), but they could not determine which elements of instruction lead to the best outcome.

Therefore, it is investigated whether certain theoretically discriminated types of instruction for writing-to-learn procure the best results. In the selection of studies, the aim was to provide as broad an overview as possible of research into instruction of writing-to-learn in various disciplines and types of education. Studies with different designs were included in the literature search: experimental, quasi-experimental and case studies. From grade 5 onwards, students appear to be capable of reflection on the texts they write (Van Gelderen, 1997). A selection criterion was therefore that studies had been carried out from grade 5 onwards in primary education, in secondary education and higher education. Another selection criterion was that studies were embedded in regular educational contexts. Only in this condition it is possible to study whether writing-to-learn leads to effects in 'real' classrooms. Based on Klein's (1999) hypotheses about cognitive processes in writing-to-learn, three types of instruction have been distinguished: Forward Search, Backward Search and Genre Writing.

Forward Search instructs students to write down everything they know about a topic without planning. Then they are instructed to reread and rewrite their texts. The assumption is that students who reread and revise their drafts will reflect on the contents and observe relations they may have not noticed before. In this way they may gain new topic knowledge and insights.

Backward Search instructs writers to first plan the text, then write, reread the result and revise the text if necessary. When students plan, they reflect on the rhetorical goal and structure of their text. When rereading and revising, they reread their planning to check whether the text aligns with the planning. If that is not the case, they must decide to adjust the text or the planning. This requires them to think about the content. This reflection can lead to new topic knowledge and insights.

Genre Writing is instruction for writing in a specific genre, such as an argument. Each genre is characterized by one or more conceptual relationships (such as cause and effect) and by a rhetorical purpose (such as informing). Using a model text, the teacher explains how the conceptual relations in the text are realized. By reflecting during writing in the relevant genre on how concepts are related and how the relations are connected to the rhetorical goal, students may gain new insights and new knowledge.

The analysis of the selected studies disclosed a fourth type of instruction: Planning Only. This instruction is only aimed at the planning of a text. Students are required to reflect on the selection of the content and the organization of the text in relation to the goal. Their reflection may lead to learning.

The instructions from the collected studies have been classified into the four distinguished types. The results of the four types were compared to each other. To that end, only studies comparing experimental conditions to control conditions were used,

because the results thereof are better comparable than results from studies without a control group. All four types often appeared to lead to effects on learning, but evidence for studies based on Genre Writing was strongest.

Four quasi-experiments

As a follow-up to the literature study, two quasi-experimental studies into the effects of Genre Writing instruction in biology and mathematics teacher education have been conducted (Chapter 3). Subsequently, two think aloud studies were conducted to determine whether indicators of the process of writing-to-learn could be identified. Additionally, two quasi-experimental studies were carried out in pre-vocational education, also in biology and mathematics (chapter 4), and followed by two think aloud studies into the process of writing-to- learn.

The quasi-experiments are studies into the effects of Genre Writing supplemented by Planning and Revising (GWPR instruction). This variant was chosen, because the literature study showed that it led to the most evidence for effectiveness. The students were asked to write for an audience that is not familiar with the subject of the text, for example students who have not yet taken the course. This requires writers to 'translate' academic language into everyday language. This is an extra stimulus for students to think about the formulation of their text. Peer response was used to stimulate revision of the text, which entails that students comment each other's draft in pairs on the way they formulated the conceptual relations.

First, a quasi-experiment was carried out in biology teacher education and a second one in mathematics teacher education (both in chapter 3). The design of the two studies is similar. The participants in biology teacher training were third-year students in the course 'History of Biology', and in mathematics the participants were third-year students in the course 'Rows and Limits'. The researcher collaborated with the teachers in developing model texts and writing assignments appropriate to the goals of the courses. Subsequently, lessons incorporating the model texts and the assignments were carried out in the experimental groups of both types of teacher education,. The control groups followed the usual program of the courses (without writing assignments). The courses were concluded with a final test, which measured insight in and knowledge of subject matter.

The other two quasi-experiments took place in pre-vocational education, one in grade 7 biology lessons and the second in grade 10 mathematics lessons (Chapter 4). The design of the studies was similar to that in teacher education. In biology, the students followed lessons on the theme 'The human body' and in mathematics on 'Relations'.

The teachers were fourth-year trainees who had participated as experimental students in the quasi-experiment in the biology teacher training course or in the mathematics teacher training course in the previous year. As a result, these trainee teachers were already familiar with writing-to-learn. As in the teacher training courses, we collaborated with these teachers on the formulation of appropriate model texts and writing assignments for the lesson themes. The experimental groups followed the lessons with model texts and writing assignments, while the control groups received regular lessons without writing assignments. In both conditions, the

lesson series were completed with a final test, aimed at insight into and knowledge of the taught subject matter.

Four think aloud studies

Following the quasi-experiments in teacher education and in pre-vocational education, think aloud studies were conducted with a selection of participants from the experimental and control groups. The aim was to investigate whether indicators of the process of writing-to- learn could be observed more often in experimental students than in control students.

The studies were conducted in individual sessions that were videotaped. The biology- and math students completed a writing assignment at the end of the lesson series. The selected students completed that writing assignment thinking aloud. The filmed sessions were transcribed and analyzed. For the analysis, a coding system was developed based on the cognitive model of the writing process of Hayes and Flower (1980). Special attention was given to seven codes related to reflection on one's own writing process, because they can be regarded as indicators for the process of writing-to-learn.

Results

The results achieved by students in teacher education show that students in both experimental conditions gained more insight into subject matter than students in the control groups. The knowledge of subject matter that the experimental biology students have acquired is also larger than that of control students. As for the mathematics students it is not clear whether the knowledge of students in the experimental condition is larger than that of control students, because the test for measuring the knowledge of mathematics students turned out to be unreliable.

The results of the quasi-experiment in grade 7 biology class of pre-vocational education do not show a difference between the conditions in acquired insight into and knowledge of subject matter. The experimental students in grade 10 mathematics class of pre-vocational education do have higher scores for insight into subject matter than students in the control condition. Differences in knowledge of subject matter could not be determined because the schools had the policy not to include multiple choice questions in math tests.

The think aloud studies demonstrated that the experimental students in biology teacher training showed three out of seven indicators of the process of writing-to-learn more often than the control group, that is: revision during the formulation of an expression, revision after writing an expression, and rereading one's own text. In mathematics teacher training, the think aloud studies demonstrated that experimental students showed four (out of seven) indicators of the process of writing-to-learn more often than control students, that is: using knowledge of the intended audience, thinking about selecting the content of the text, revising during the formulation of an utterance and rereading one's own text.

The results of the think aloud studies in pre-vocational education show a different picture. No differences were found between the two conditions for grade 7 students who followed the biology lessons. Among grade 10 students who followed the experimental mathematics lessons, more indicators were found than among students in the control group. These are the following indicators: using knowledge of

the intended audience, revising after formulating an expression, rereading one's own text and reconsidering the task approach.

Conclusions

The expectation was that GWPR instruction could lead to new insights into and new knowledge of subject matter for the students. This was the case in biology teacher education. In mathematics teacher education, students gained new insights into subject matter, but differences in topic knowledge could not be measured because of an unreliable test. In grade 10 mathematics classes of pre-vocational education, GWPR instruction led to new insights into subject matter. New topic knowledge could not be measured, because schools did not allow multiple choice tests in the discipline mathematics. In grade 7 biology classes of pre-vocational education, no effect of GWPR instruction on insight or topic knowledge could be demonstrated. This result can be attributed to the fact that for students in biology lessons available lesson time was halved; therefore, they were at a disadvantage compared to students in mathematics lessons.

The conclusion is that GWPR instruction in writing-to-learn can be an instrument to support students in teacher education to better understand subject matter. This also applies to grade 10 students in mathematics classes in pre-vocational education. For students in grade 7 biology class in pre-vocational education, writing-to-learn may also have positive effects on their topic knowledge of and insight into subject matter, on the condition that there is considerably more time for implementation, because time is an important factor, as Hand et al. (2021) show.

Discussion

The effect of GWPR instruction can be explained by the fact that it initiates a coherent writing process by repeatedly directing students' and learners' attention to conceptual relationships characteristic of the genre being written, and to the intended audience that is not familiar with the topic. In this way, students are constantly stimulated to think about how to realize and formulate conceptual relationships. They also have to think repeatedly about the question whether their text is understandable to their audience. This frequent focus on conceptual relationships and the intended audience seems to bring about the effects.

GWPR instruction did not lead to new insights and new topic knowledge for grade 7 students in biology. The explanation for this is that the students in the experimental condition performed a complete writing assignment in only one lesson (45 minutes), while for students in the other three quasi-experiments two lessons were available for the performance of each writing assignment. One lesson is used for the first part of the writing assignment (planning and writing a first draft) and the second lesson for the second part of the writing assignment (commenting and receiving comments, rereading and revising the text). The grade 7 students in biology probably had too little time to familiarize with writing assignments that were new to them. In addition, they probably didn't get enough opportunity to step away from their own text for meaningful revision, because everything had to be done in the same lesson. So, more time was needed to master executing writing assignments (see also Hand et al., 2021).

The expectation in the four think aloud studies was that students from the experimental groups would use more indicators for the process of writing-to-learn

than students from the control groups. Three of the think aloud studies (two in teacher training and one in grade 10 pre-vocational education) do indeed show more indicators in the experimental condition than in the control condition. However, no differences were found between the two conditions in seventh grade students who took biology lessons. These outcomes converge with the results of the final tests found in the four quasi-experiments. The halving of the biology lesson time among students in grade 7 apparently resulted not only in the absence of effects in terms of insight into and topic knowledge of subject matter, but also in the absence of indicators for writing-to-learn in the think aloud study in question.

Among students of biology and mathematics teacher education, there were some notable differences between the observed indicators. These differences seem to be due to differences in the writing assignments that students received. The biology students' assignment did not state that they should target a specific audience. As a result, the code 'use knowledge about the public' did not appear among the biology students, but only among the mathematics students who were given that assignment. A second difference was that the biology students were given a writing assignment stating where they could find information about the subject of their text, while this was not the case for the mathematics students. As a result, the biology students did not think about the selection of the content of the text while the mathematics students did.

Limitation

A limitation of the quasi-experimental studies is that the samples are quite small. With larger samples, the results would be more generalizable. The limited sample size in teacher training is due to the fact that in the studied academic years and types of education only one class per academic year joined the series of lessons. In pre-vocational education, the availability of trained teachers (student teachers who had participated in the writing-to-learn experiments in the previous year) was limited. The think aloud studies involved very small samples. Therefore, our results need to be supplemented with findings from studies with larger sample sizes to determine whether the results are generalizable.

Recommendations

In the studies discussed in this thesis a truly experimental design is preferred; however, the situation in the institutions involved was not such that it was possible to use it. We therefore recommend conducting experimental studies in the future. This can be done by randomly assigning students to one of two conditions in one class. The teacher instructs half of the students to write about topic A and the other half about topic B. In this way, the two groups of students can function as each other's control group.

The literature review found empirical evidence for the effectiveness of four types of instruction in writing-to-learn. However, these results are based on small numbers of (quasi) experimental studies. Only the effects of Genre Writing are based on a larger number of studies, that is eleven. It is possible that a particular type of instruction can only lead to positive effects in certain contexts, depending on the type of students. To gain more certainty about which type of instruction in writing-to-learn

is effective in which contexts or disciplines, it is important that future studies focus on these questions.

To investigate the process of writing-to-learn, it was decided to carry out think aloud studies. It is possible that reflection moments (indicators) of students have been missed during the implementation. One method to cope with this is to use keystroke logging (Baaijen, 2012) in addition to think aloud studies. With keystroke logging, all students' keystrokes are registered and pauses in thinking can be accurately registered as well.

Implications for education

So far, writing-to-learn has been little used in regular education. This has to do with the fact that subject teachers are reluctant to use writing assignments in their lessons (Baker, 2008). It is therefore recommended to provide training to subject teachers interested in using writing-to-learn. In a course, they can work under the supervision of an expert on composing and embedding writing assignments in a series of lessons that they regularly teach. Group work is suitable for adapting writing assignments to the lesson objectives. Sharing ideas gives teachers confidence in their ability to tailor writing assignments to writing-to-learn by students and develop model texts to familiarize students with genre-specific features of texts. In this thesis, six activities are described as components of such training.

Writing-to-learn can be applied as part of language policy. Language policy aims to familiarize students and pupils with academic language use in various subjects, so that they are able to understand lesson content. In practice, this means that subject teachers not only pay attention to the subject matter in their lessons, but also to the language of the subject. The use of writing-to-learn can be a good fit with this, because it requires students to write about the subject in everyday language. The 'translation' of academic language into everyday language by students can deepen their understanding of the language. At the same time, students can gain more insight into and topic knowledge of subject matter.

Samenvatting

Effecten van instructie in schrijvend leren in verschillende vakken en vormen van onderwijs

Studenten in alle vormen van onderwijs hebben vaak moeite met het verwerven van vakkennis en inzicht in de lesstof van verschillende vakken. Ze begrijpen theorie uit hun lesboek niet of ze kunnen die niet onthouden (Sampson & Phelps Walker, 2012; Hunter & Tse, 2013; Finkenstaedt-Quinn et al., 2017). Gewoonlijk leren studenten theorie door studieboeken te bestuderen, door goed te luisteren naar de uitleg van hun docent of door lesstof samen met andere studenten te bespreken. Toch is dat vaak niet voldoende om greep op de lesstof te krijgen. Schrijvend leren kan daarbij helpen, maar wordt nog weinig toegepast in het onderwijs. Bij schrijvend leren is het de bedoeling om het leren te bevorderen door teksten te schrijven over de lesstof. Hierdoor kunnen studenten het geschrevene teruglezen en erover nadenken of ze de vakinhoud goed weergegeven hebben. Schrijvend leren is schrijven met het doel om inzicht in lesstof te krijgen en vakkennis op te doen.

Schrijvend leren wordt al sinds de jaren 80 van de vorige eeuw onderzocht. Aanvankelijk werd gedacht dat het schrijven op zich het leerproces op gang brengt (strong text theory), maar dat bleek lang niet altijd het geval te zijn (Klein, 1999). Een noodzakelijke voorwaarde om het denken van studenten te stimuleren is instructie van cognitieve en metacognitieve strategieën bij schrijftaken, leidend tot nieuwe inzichten in en kennis van de lesstof (Bangert-Drowns et al., 2004). Een voorbeeld van instructie van een cognitieve strategie bij een schrijftaak is: ‘probeer eerst de inhoud van de tekst die je gaat schrijven op een zinvolle manier te organiseren’. Een voorbeeld van instructie van een metacognitieve strategie is: ‘schrijf op welke onderwerpen uit de les je nog niet begrijpt’ (Nückles et al., 2009). De instructies zetten studenten en leerlingen aan tot denken. Denken kan ze helpen om tot nieuwe inzichten in en kennis van de lesstof te komen (Bereiter & Scardamalia, 1987).

Dit proefschrift begint met een literatuurstudie door middel waarvan is onderzocht welke typen instructie bij schrijftaken positieve effecten hebben op het leren in verschillende onderwijstypen en vakken. In vervolg daarop zijn vier quasi-experimentele studies uitgevoerd naar de effecten van een bepaald type instructie in verschillende onderwijstypen en vakken. In aansluiting op elk van de vier quasi-experimenten is een hardopdenkstudie gedaan naar het proces van schrijvend leren. In totaal zijn negen studies uitgevoerd, die hieronder worden besproken.

Literatuuronderzoek

Het doel van het literatuuronderzoek in dit proefschrift (hoofdstuk 2) is om een overzicht te geven van het onderzoek naar effecten van verschillende typen instructie

in schrijvend leren. De suggestie van Bangert-Drowns et al. (2004) om te onderzoeken of en hoe instructie van (meta)cognitieve strategieën bij schrijftaken kan leiden tot nieuwe inzichten en nieuwe vakkennis is in een groot aantal onderzoeken ter harte genomen. In enkele overzichtsstudies worden positieve effecten van schrijvend leerinstructies op leerprocessen gerapporteerd (Miller et al., 2018; Graham et al., 2020; Hand et al., 2021), maar er kon niet worden vastgesteld welke elementen van instructie tot het beste resultaat leiden.

Daarom is in dit hoofdstuk onderzocht of bepaalde theoretisch onderscheiden typen instructie voor schrijvend leren de beste resultaten opleveren. Bij de selectie van studies werd gestreefd naar een zo breed mogelijk overzicht van onderzoek naar instructie van schrijvend leren in verschillende vakken en typen onderwijs. Ook zijn onderzoeken met verschillende designs in het literatuuronderzoek opgenomen: experimentele, quasi-experimentele en case studies. Vanaf groep 7 blijken leerlingen goed in staat te zijn om na te denken over de teksten die ze schrijven (Van Gelderen, 1997). Een selectiecriterium was daarom dat studies waren uitgevoerd vanaf groep 7 in het basisonderwijs, in het voortgezet onderwijs en in het hoger onderwijs. Een ander selectiecriterium was dat studies uitgevoerd waren in reguliere onderwijscontexten. Alleen dan is het mogelijk om te weten te komen of schrijvend leren werkt in 'echte' onderwijssituaties. Op basis van hypotheses van Klein (1999) over cognitieve processen bij schrijvend leren, zijn drie typen instructie onderscheiden: Forward Search, Backward Search en Genre Writing.

Forward Search instrueert studenten om zonder plan alles wat ze weten over een onderwerp op te schrijven. Daarna worden ze geïnstrueerd hun teksten terug te lezen en te herschrijven. De veronderstelling is dat studenten die hun klad herlezen en reviseren gaan nadenken over de inhoud en dat ze verbanden zien die ze nog niet eerder opgemerkt hebben. Zo kunnen ze tot nieuwe vakkennis en inzichten komen.

Backward Search instrueert schrijvers eerst een plan voor de tekst te maken, vervolgens te schrijven, het resultaat daarvan te herlezen en de tekst zo nodig te reviseren. Als studenten plannen, denken ze na over het doel en een structuur voor hun tekst. Bij het herlezen en reviseren moeten ze hun planning teruglezen om na te gaan of de tekst aansluit bij de planning. Als dat niet het geval is moeten ze besluiten om de tekst of de planning aan te passen. Dat vereist dat ze nadenken over de inhoud. Deze reflectie kan leiden tot nieuwe vakkennis en inzichten.

Genre Writing is instructie voor het schrijven in een bepaald genre, bijvoorbeeld een betoog. Elk genre wordt gekenmerkt door een of meer conceptuele relaties (zoals oorzaak-gevolg) en door een retorisch doel (zoals informeren). Aan de hand van een modeltekst licht de docent toe hoe de conceptuele relaties in de tekst gerealiseerd worden. Door bij het schrijven in het betreffende genre na te denken over de manier waarop concepten verbonden worden en hoe de verbanden gerelateerd zijn aan het doel, kunnen studenten nieuwe inzichten en nieuwe kennis opdoen.

Bij de analyse van de geselecteerde studies kwam een vierde instructietype naar voren: Planning Only. Deze instructie is alleen gericht op de planning van een tekst. Studenten moeten nadenken over de selectie van de inhoud en de organisatie van de tekst in relatie tot het doel. Hun reflectie hierover kan tot leren leiden.

De gehanteerde instructies uit de verzamelde studies zijn ingedeeld in de vier onderscheiden typen. De resultaten van de vier typen zijn met elkaar vergeleken. Daarvoor zijn alleen studies gebruikt die experimentele condities met een

controleconditie vergeleken, omdat de resultaten daarvan beter vergelijkbaar zijn dan resultaten uit studies zonder controlegroep. Alle vier de typen bleken vaak effect op het leren te hebben, maar het bewijs voor studies gebaseerd op Genre Writing was het sterkst.

Vier quasi-experimenten

In vervolg op de literatuurstudie zijn twee quasi-experimentele studies gedaan naar de effecten van Genre Writing instructie in de tweedegraads lerarenopleidingen voor biologie en wiskunde (hoofdstuk 3). Aansluitend daarop zijn twee hardopdenkstudies uitgevoerd om na te gaan of indicatoren van het proces van schrijvend leren vastgesteld konden worden. Vervolgens zijn twee andere quasi-experimentele studies in het vmbo uitgevoerd, eveneens bij biologie en wiskunde (hoofdstuk 4), en ook gevolgd door twee hardopdenkstudies naar het proces van schrijvend leren.

De quasi-experimenten zijn studies naar effecten van Genre Writing aangevuld met Planning and Revising (GWPR-instructie). Deze variant is gekozen, omdat uit de literatuurstudie bleek dat hiervoor het meeste bewijs van effectiviteit bestaat. Aan de studenten is gevraagd om te schrijven voor een publiek dat niet vertrouwd is met het onderwerp van de tekst, bijvoorbeeld studenten die het vak nog niet gevolgd hebben. Dat vraagt van schrijvers om een ‘vertaling’ te maken van academisch naar dagelijks taalgebruik. Het is voor studenten een extra stimulans om na te denken over de formulering van hun tekst. Voor het stimuleren van revisie van de tekst is gebruik gemaakt van ‘peer response’ waarbij studenten elkaar in duo’s commentaar geven op de manier waarop ze de conceptuele relaties hebben geformuleerd in de eerste versie van hun tekst.

Allereerst is een quasi-experiment uitgevoerd in de lerarenopleiding biologie en een tweede in de lerarenopleiding wiskunde (beide in hoofdstuk 3). De opzet van de twee studies is dezelfde. De deelnemers bij biologie waren derdejaars studenten in de cursus ‘Geschiedenis van de biologie’ en bij wiskunde betrof het derdejaars studenten in de cursus ‘Rijen en limieten’. Er werd samengewerkt met de docenten bij het ontwikkelen van passende modelteksten en schrijfopdrachten bij de doelen van de cursussen. Vervolgens werden in de experimentele groepen van beide opleidingen lessen uitgevoerd waarin de modelteksten en de opdrachten verwerkt waren. De controlegroepen volgden het gebruikelijke programma van de cursussen (zonder schrijfopdrachten). De cursussen werden afgesloten met een eindtoets, waarmee inzicht en kennis van de lesstof werden gemeten.

De andere twee quasi-experimenten vonden plaats in het vmbo, een in biologielessen in de brugklas mavo en een tweede in wiskundelessen in klas 4 van de kaderberoepsgerichte opleiding (hoofdstuk 4). De opzet van de onderzoeken was dezelfde als die in de lerarenopleidingen. Bij biologie volgden de leerlingen lessen over het thema ‘Het menselijk lichaam’ en bij wiskunde over ‘Verbanden’. De docenten waren vierdejaars stagiaires die in het voorafgaande jaar als experimentele student deel hadden genomen aan het quasi-experiment in de lerarenopleiding biologie of in de lerarenopleiding wiskunde. Daardoor waren deze stagedocenten al vertrouwd met schrijvend leren. Net als in de lerarenopleidingen is er samengewerkt met deze docenten aan de formulering van passende modelteksten en schrijfopdrachten bij de lesthema’s. De experimentele groepen volgden de lessen met modelteksten en schrijfopdrachten terwijl de controlegroepen reguliere lessen kregen

zonder schrijfopdrachten. In beide condities werden de lessenseries afgerond met een eindtoets, gericht op inzicht in de lesstof.

Vier hardopdenkstudies

Aansluitend op de quasi-experimenten in de lerarenopleidingen en in het vmbo, zijn hardopdenkstudies uitgevoerd met een selectie van de deelnemers in de experimentele en controlegroepen. Doel was om te onderzoeken of indicatoren van het proces van schrijvend leren vaker vastgesteld konden worden bij experimentele studenten dan bij controlestudenten.

De studies werden uitgevoerd in individuele sessies die gefilmd werden. De biologie- en wiskundestudenten maakten een schrijfopdracht aan het eind van de lessenserie. De geselecteerde studenten maakten die schrijfopdracht hardop denkend. De gefilmde sessies werden getranscribeerd en geanalyseerd. Voor de analyse is een coderingssysteem ontwikkeld op basis van het cognitieve model van het schrijfproces van Hayes en Flower (1980). Speciale aandacht kregen zeven codes die betrekking hebben op reflectie op het eigen schrijfproces, omdat deze beschouwd kunnen worden als indicatoren voor het proces van schrijvend leren.

Resultaten

De resultaten die studenten van de lerarenopleidingen behaalden, laten zien dat studenten in beide experimentele condities meer inzicht in de lesstof hebben verkregen dan de studenten in de controlegroepen. De kennis van de lesstof die de experimentele biologiestudenten hebben gekregen is ook groter dan die van controlestudenten. Voor de wiskundestudenten is niet duidelijk of de kennis van studenten in de experimentele conditie groter is dan die van controlestudenten, omdat de toets voor het meten van kennis van wiskundestudenten niet betrouwbaar bleek.

De resultaten van het quasi-experiment in de mavo brugklas bij biologie laten geen verschil zien tussen de condities in verworven inzicht in en kennis van de lesstof. De experimentele wiskundestudenten in 4-vmbo-kader hebben wel hogere scores op inzicht in de lesstof dan studenten in de controleconditie. Verschillen in kennis van de lesstof konden niet bepaald worden, omdat de scholen het beleid hadden om in wiskundetoetsen geen meerkeuzevragen te stellen. Uit de hardopdenkstudies bleek dat de experimentele studenten in de lerarenopleiding biologie in het voordeel waren voor drie (van de zeven) indicatoren: reviseren gedurende het formuleren van een uiting, reviseren na het schrijven van een uiting en herlezen van de eigen tekst. In de lerarenopleiding wiskunde waren de experimentele studenten in het voordeel bij vier (van de zeven) indicatoren: kennis van het beoogde publiek gebruiken, nadenken over het selecteren van de inhoud van de tekst, reviseren gedurende het formuleren van een uiting en herlezen van de eigen tekst.

De resultaten van de hardopdenkstudies in het vmbo laten een ander beeld zien. Voor brugklasleerlingen die het vak biologie volgden, zijn geen verschillen gevonden tussen de beide condities. Bij de 4-vmbo-kaderleerlingen die de experimentele lessen wiskunde hebben gevolgd zijn meer indicatoren gevonden dan bij leerlingen in de controlegroep. Dat zijn de volgende: kennis van het beoogde publiek gebruiken, reviseren na het formuleren van een uiting, herlezen van de eigen tekst en heroverwegen van de taakaanpak.

Conclusies

De verwachting was dat GWPR-instructie kan leiden tot nieuwe inzichten in en nieuwe kennis van lesstof bij de studenten. Dat was in de lerarenopleiding biologie het geval. In de lerarenopleiding wiskunde kwamen studenten tot nieuwe inzichten in de lesstof, maar konden verschillen in kennis niet gemeten worden door een onbetrouwbare toets. In het vmbo leidde GWPR-instructie tot nieuwe inzichten voor het vak wiskunde. Nieuwe kennis kon niet gemeten worden omdat scholen geen meerkeuzetoetsen toelieten. Bij vmbo biologie kon geen effect van GWPR instructie op inzicht of kennis aangetoond worden. De verklaring is dat de beschikbare lestijd gehalveerd was voor leerlingen in de biologielessen en dat zij daardoor in het nadeel waren ten opzichte van de leerlingen in de wiskundelessen. De conclusie is dat GWPR-instructie bij schrijvend leren een middel kan zijn om studenten in de lerarenopleidingen te helpen de lesstof beter te begrijpen. Dit geldt ook voor leerlingen in wiskundelessen in het vmbo. Voor leerlingen in de brugklas van het vmbo heeft schrijvend leren in biologielessen misschien ook positieve effecten op hun vakkennis en inzichten in de lesstof, als er aanzienlijk meer tijd voor de uitvoering is, want tijd is een belangrijke factor, zoals Hand et al. (2021) laten zien.

Discussie

Het effect van GWPR-instructie kan worden verklaard vanuit het feit dat deze een coherent schrijfproces op gang brengt door de aandacht van studenten en leerlingen herhaaldelijk te richten op conceptuele relaties die kenmerkend zijn voor het genre waarin geschreven wordt en op het beoogde publiek dat niet vertrouwd is met het onderwerp. Zo worden studenten en leerlingen telkens weer gestimuleerd om na te denken over de manier waarop zij conceptuele relaties kunnen realiseren en formuleren. Ook moeten zij herhaaldelijk nadenken over de vraag of hun tekst begrijpelijk is voor hun publiek. Die herhaling van de aandachtspunten lijkt de effecten tot stand te brengen.

Bij brugklasleerlingen in het vak biologie leidde GWPR-instructie niet tot nieuwe inzichten en nieuwe kennis. De verklaring daarvoor is dat de leerlingen in de experimentele conditie een complete schrijfopdracht in één les (45 minuten) uitvoerden, terwijl voor studenten en leerlingen in de andere drie quasi-experimenten twee lessen beschikbaar waren voor de uitvoering van elke schrijfopdracht. Eén les werd gebruikt voor het uitvoeren van het eerste deel van de schrijfopdracht (planning en het schrijven van een eerste versie) en de tweede les voor het tweede deel van de schrijfopdracht (commentaar krijgen, herlezen en herschrijven van de tekst). De brugklasleerlingen hadden vermoedelijk te weinig tijd om vertrouwd te raken met schrijfopdrachten die nieuw voor hen waren. Bovendien kregen ze waarschijnlijk onvoldoende gelegenheid om afstand te nemen van hun eigen tekst voor zinvolle revisie, omdat alles in dezelfde les moest gebeuren. Er was dus meer tijd nodig om het werken met schrijvend leren opdrachten onder de knie te krijgen (zie ook Hand et al., 2021).

De verwachting bij de vier hardopdenkstudies was dat meer indicatoren voor het proces van schrijvend leren gevonden zouden worden bij studenten/leerlingen uit de experimentele dan uit de controlegroepen. Drie van de hardopdenkstudies (bij de lerarenopleidingen en 4-vmbo-kader) laten inderdaad meer indicatoren zien bij de experimentele conditie dan bij de controleconditie. Maar bij brugklasleerlingen die

lessen biologie volgden zijn geen verschillen tussen de twee condities gevonden. Deze uitkomsten convergeren met de gevonden resultaten van de eindtoetsen in de vier quasi-experimenten, De halvering van de lestijd biologie bij de leerlingen in de brugklas heeft vermoedelijk niet alleen geresulteerd in de afwezigheid van effecten qua kennis en inzicht in de lesstof maar ook in de afwezigheid van indicatoren voor schrijvend leren in de betreffende hardopdenkstudie.

Bij studenten van de lerarenopleidingen biologie en wiskunde waren er enkele opmerkelijke verschillen tussen de geobserveerde indicatoren. Die verschillen lijken het gevolg te zijn van verschillen in de schrijfopdrachten die studenten kregen. In de opdracht van de biologiestudenten stond niet dat ze zich tot een specifiek publiek moesten richten. Daardoor kwam de code 'kennis over het publiek gebruiken' niet voor bij de biologiestudenten en wel bij de wiskundestudenten die die opdracht wel kregen. Een tweede verschil was dat de biologiestudenten een schrijfopdracht kregen waarin vermeld stond waar ze informatie over het onderwerp van hun tekst konden vinden, terwijl dat in de opdracht van de wiskundestudenten niet het geval was. Dit had tot gevolg dat de biologiestudenten niet nadachten over de selectie van de inhoud van de tekst en de wiskundestudenten wel.

Beperking

Een beperking van de experimentele studies is dat de steekproeven vrij klein zijn. Met grotere steekproeven zouden de uitkomsten beter generaliseerbaar zijn. De beperkte steekproefgrootte in de lerarenopleiding heeft te maken met het feit dat in de onderzochte schooljaren en opleidingen slechts één groep per jaar de lessenserie volgde. In het vmbo was de beschikbaarheid van getrainde docenten (leraren in opleiding die hadden deelgenomen aan de schrijvend leren experimenten in het jaar ervoor) beperkt. Voor de hardopdenkstudies geldt dat er sprake was van zeer kleine steekproeven. Deze studies behoeven aanvulling met uitkomstenvan studies met grotere steekproeven om te bepalen of de resultaten generaliseerbaar zijn.

Aanbevelingen

In dit proefschrift zijn vier quasi-experimentele studies besproken. Een echt experimenteel design heeft de voorkeur, maar de situatie in de betrokken instellingen leende zich daar niet voor. We doen daarom de aanbeveling om in de toekomst experimentele studies uit te voeren. Dat kan door in een klas studenten willekeurig aan een van beide condities toe te wijzen. De docent instrueert de ene helft van de groep om te schrijven over onderwerp A en de andere helft over onderwerp B. Op deze manier zijn de twee groepen studenten elkaars controlegroep.

In de literatuurstudie is empirisch bewijs gevonden voor de effectiviteit van vier typen instructie in schrijvend leren. Deze resultaten zijn echter gebaseerd op kleine aantallen (quasi) experimentele studies. Alleen de effecten van Genre Writing zijn gebaseerd op een groter aantal studies, namelijk elf. Het is mogelijk dat een bepaald type instructie alleen tot positieve effecten kan leiden in bepaalde contexten, afhankelijk van het type studenten. Om meer zekerheid te krijgen over de vraag welk type instructie effectief is in welke contexten of disciplines, is het belangrijk dat toekomstige studies gericht zijn op deze vragen.

Voor het onderzoeken van het proces van schrijvend leren is gekozen voor het uitvoeren van hardopdenkstudies. Het kan zijn dat bij de uitvoering daarvan

reflectiemomenten (indicatoren) van studenten gemist zijn. Een methode om hieraan het hoofd te bieden is het gebruik van keystroke logging (Baaijen, 2012) naast hardop denkstudies. Bij keystroke logging worden alle toetsaanslagen van studenten geregistreerd en kunnen denkpauzes nauwkeurig geregistreerd worden.

Implicaties voor het onderwijs

Schrijvend leren wordt tot nu toe weinig gebruikt in het reguliere onderwijs. Dat heeft te maken met het feit dat vakdocenten er tegenop zien om schrijfopdrachten in hun lessen te gebruiken (Baker, 2008). Daarom wordt aanbevolen om vakdocenten die geïnteresseerd zijn in het gebruiken van schrijvend leren een training te geven. In een training kunnen zij onder begeleiding van een expert werken aan het samenstellen en inbedden van schrijfopdrachten in een lessenserie die zij regelmatig geven. Hierbij is groepswerk geschikt om schrijfopdrachten passend te maken bij de lesdoelstellingen. Het uitwisselen van ideeën geeft docenten vertrouwen in hun vermogen om schrijfopdrachten geschikt te maken voor schrijvend leren van leerlingen en modelteksten te ontwikkelen om leerlingen vertrouwd te maken met genrespecifieke kenmerken van teksten. In dit proefschrift zijn zes activiteiten beschreven als onderdelen van een dergelijke training.

Schrijvend leren kan toegepast worden als onderdeel van taalbeleid. Taalbeleid heeft tot doel om studenten en leerlingen vertrouwd te maken met academisch taalgebruik in verschillende vakken, zodat ze in staat zijn om lesinhouden te begrijpen. In de praktijk betekent dit dat vakdocenten in hun lessen niet alleen aandacht besteden aan de lesstof, maar ook aan de taal van het vak. Het gebruik van schrijvend leren kan daar goed op aansluiten, omdat dit van studenten en leerlingen vraagt om over het vak in alledaagse taal te schrijven. Het ‘vertalen’ van academische taal in alledaagse taal door studenten en leerlingen, kan hun begrip van de taal verdiepen. Tegelijkertijd kunnen studenten en leerlingen meer inzicht in en kennis van de lesstof verwerven.

About the author

Education
Aartje van Dijk was born on February 10, 1955, in Zeist. After obtaining the certificate from the college preparatory department (in Dutch: a gymnasium) of a combined school called Christelijk Lyceum (in Zeist), she entered Utrecht University to study Dutch language and literature from 1973 to 1977. During the study she became interested in general linguistics and joined a master about that field after finishing the bachelor study (1977-1980). In addition, she obtained her teaching qualification for Dutch in secondary education (1978). In that period, she started as a volunteer to teach Dutch to adult Moroccan migrants (1977-1987). Therefore, second language acquisition was the element of linguistics that particularly attracted her to specialize in. Her master thesis conducted at the Utrecht Institute for Developmental Psychology was directed at the question how the second language acquisition process takes place (1980).

Employment
She kept focusing on Dutch as a second language in various jobs following her graduation in general linguistics. She collaborated with volunteers in community centers, first in Almelo (De Roggekamp) and later in Utrecht (Educatief Centrum) to supervise them in their lessons Dutch as a second language for adult migrants (1980-1987). However, when education in Dutch as a second language developed into "basic adult education" (basiseducatie), other tasks were provided to her. Still employed at the Educatief Centrum in Utrecht, she collaborated with teachers in bridging programs and in vocational education for adult students to implement a content-based approach in their classes (1987-1994). The content based approach builds on the teaching of Dutch as a second language and supports students to learn disciplinary and academic language

Changing from adults to secondary school students
In 1994, Aartje started at the CED-Group in Rotterdam, a school advisory service. She became a member of a project-team directed at teachers of secondary school-students learning Dutch as a second language (in Dutch: Internationale Schakelklassen, ISK). The project team and the teachers worked together and determined which facilities were needed to arrive at students' sufficient acquisition of Dutch. The team composed teaching materials, curricula etc. In 2001 she broadened her field of activities to students in secondary schools by participating in various projects, for instance an implementation project of language policy in Rotterdam schools (2001-2006).

Teacher education and Phd
After a short employment at a private school organizing courses Dutch as a second language (2006-2008), she started at Rotterdam University of Applied Sciences as a teacher educator of Dutch as a second language and Dutch linguistics. In 2012, she started the PHD study about writing-to-learn.