frontiers | Frontiers in Education

TYPE Original Research
PUBLISHED 07 July 2023
DOI 10.3389/feduc.2023.1094156

OPEN ACCESS

EDITED BY
Cheryl J. Craig,
Texas A and M University, United States

REVIEWED BY
Abdullah Kaldırım,
Dumlupinar University, Türkiye
Jesse R. Sparks,
Educational Testing Service, United States

*CORRESPONDENCE
Aartje Van Dijk
a.c.van.dijk@hr.nl

†These authors share senior authorship

RECEIVED 09 November 2022
ACCEPTED 15 June 2023
PUBLISHED 07 July 2023

CITATION
Van Dijk A, Van Gelderen A and Kuiken F (2023) Writing-to-learn in biology and mathematics teacher education: promoting students' topic knowledge and insight.
Front. Educ. 8:1094156.
doi: 10.3389/feduc.2023.1094156

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

Aartje Van Dijk[1]*, Amos Van Gelderen[1†] and Folkert Kuiken[2†]

[1]Research Centre for Urban Talent, Rotterdam University of Applied Sciences, Rotterdam, Netherlands, [2]Faculty of Humanities, Amsterdam Center for Language and Communication, University of Amsterdam, Amsterdam, Netherlands

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. Finally, we discuss the impact of the results for the theory, follow-up studies and teaching practice.

KEYWORDS

writing-to-learn, genre writing instruction in learning, teaching cognitive strategies, reflection on own learning, co-creating a writing task

1. Introduction

Students in higher education often have difficulty understanding the contents of their textbooks and their teacher educators' explanations of subject matter (Hunter and 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 and Tse, 2013).

Several studies emphasize the importance of supporting students in acquiring the desired topic knowledge and insight (Sampson and Phelps Walker, 2012; Hunter and 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 and Carotti, 2003; Hand et al., 2009; Klein and 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 (Bereiter and Scardamalia, 1987; Galbraith, 2009). 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 and 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 in 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.

1.1. 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 and Boscolo, 2016). One often-mentioned reason is that many teachers do not know how to implement writing-to-learn in their curriculum (Akkus and Hand, 2011; Eaton and Wade, 2014; Kenney et al., 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 0.30 (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 focusing on 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 in 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. (2022) reviewed 43 studies (from grade 5 to higher education) investigating which types of instruction in 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 are 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 studies. The first type called Forward Search, stimulates the use of metacognitive strategies for reflection on contents of a draft. The second type 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 planning. The third type was called Genre Writing consisting of the provision of genre knowledge added with cognitive and/or metacognitive strategies for planning and revision. The fourth type was called Planning Only and 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 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 (Hand et al., 2004; Hohenshell and Hand, 2006; Hand et al., 2007; Finkenstaedt-Quinn et al., 2017) 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 write in, they may be able to recognize relations between concepts they were not aware of previously (Newell, 1984). Klein and Kirkpatrick (2010) and Klein and 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 and 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., 2022: 38%).

1.2. Promising elements of instruction in writing-to-learn

In this study, the evidence that the elements genre knowledge, planning and revising (cognitive and metacognitive strategies) and a lay audience provide positive effects on learning, 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., 2022). 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 and 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 into the meaning of these comparisons (Langer and 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 based on 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 in the conceptual relations and their original formulation, which may lead to new insights (Bereiter and 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 must 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.

1.3. 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 38 50-min 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 and Scardamalia (1987), Galbraith (2009) did not measure the processes involved in writing-to-learn directly. However, in an experimental study, Baaijen and 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 and 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. Thereafter, they explained its 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 and 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.

1.4. 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., 2022). 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 and 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 (Bereiter and Scardamalia, 1987; Klein, 2004; Baaijen and Galbraith, 2018).

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?

2. Study 1: writing-to-learn in biology teacher-training

2.1. Materials and methods

2.1.1. 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 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 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 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 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

2.1.2. 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.

2.1.3. 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 Zeiss (1999) 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.

2.1.4. Instruments

2.1.4.1. 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 (Reece et al., 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 0.81 (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 0.001 and for the eight items on insight 0.49. However, 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.

2.1.4.2. Vocabulary test

A vocabulary test of 30 items derived from the Dutch version of the Peabody picture vocabulary test (Dunn and 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 $p = 0.36$ and $p = 0.86$. Cronbach's alpha of the vocabulary test was 0.83, which is considered acceptable.

2.1.4.3. 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 0.53.

One open-ended question was excluded from the post-test, because of an unclear formulation. The researcher and the biology teacher educator independently coded teacher candidates' answers on the remaining open-ended questions of the post-test for a sample of eight cases. The inter-rater reliability was 0.88 (Pearson Correlation). Cronbach's alpha was 0.76 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.

2.1.4.4. 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 Darwin (2010). 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.

2.1.5. Procedure

Table 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). 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 min). A 15 min pause was inserted, but still the duration of experimental lessons was longer (some 35 min) than the control lessons. Unfortunately, we therefore cannot exclude that this difference in time-on-task has influenced the results of the experiment.

Think aloud procedures can inform about the cognitive processes involved in task execution of several sorts (Ericsson and Simon, 1993; Van Someren et al., 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 *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 1 h maximum. When teacher candidates kept silent for 10 s, the researcher encouraged them to keep thinking aloud, and used prompts such as: say aloud what you are thinking. The sessions were video recorded.

TABLE 3 Lesson structure in control and experimental condition of biology teacher education.

Control group	Experimental group	
Business as usual (lessons 2–7)	First draft (lessons 2, 4, and 6)	Revision (lessons 3, 5, and 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.

2.1.5.1. 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 and Scardamalia, 1987; Klein, 2004; Baaijen and Galbraith, 2018). 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.

2.1.6. Data analysis

2.1.6.1. 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 0.05.

2.1.6.2. Process analysis

For determining whether utterances that were considered as 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 0.20 are considered as no effect, the range between 0.20 and 0.50 is considered as a small effect, the range between 0.50 and 0.80 as a medium effect and ≥ 0.80 as a large effect.

2.2. Results

Table 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 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.

N = 38	Experimental group (*N* = 18)	Control group (*N* = 20)
Variables	Mean (SD)	Mean (SD)
Prior insight[a]	3.81 (1.66)	4.50 (1.61)
Prior topic knowledge[b]	3.83 (1.15)	4.35 (1.27)
Pre-test vocabulary[c]	18.11 (0.40)	19.70 (4.61)
Post-test insight[d]	5.28 (1.66)	3.40 (1.82)
Post-test topic knowledge[e]	13.33 (2.64)	10.85 (2.41)

[a]Theoretical maximum score: 15.5.
[b]Theoretical maximum score: 8.
[c]Theoretical maximum score: 30.
[d]Theoretical maximum score: 13.5.
[e]Theoretical maximum score: 21.

2.2.1. Prior knowledge

An ANOVA showed that experimental and control students' scores on prior insight did not differ significantly, $F(1,36) = 1.712$, $p = 0.199$, nor did the two groups differ significantly on prior topic knowledge, $F(1,36) = 0.716$, $p = 0.198$.

2.2.2. Vocabulary

An ANOVA showed no significant differences between the experimental and control teacher candidates' scores on pre-test vocabulary: $F(1,36) = 0.783$, $p = 0.382$.

2.2.3. 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, $F(1,36) = 4.479$, $p = 0.042$, partial $\eta^2 = 0.120$ (medium), as well as on post-test topic knowledge, $F(1,36) = 5.997$, $p = 0.020$, partial $\eta^2 = 0.154$ (large). The covariate prior topic knowledge did not predict differences in post-test insight, $F(1,36) = 0.218$, $p = 0.643$, nor in topic knowledge, $F(1,36) = 3.393$, $p = 0.074$. The covariate vocabulary did not predict differences in insight $F(1,36) = 0.073$, $p = 0.789$, nor on topic knowledge: $F(1,36) = 0.774$, $p = 385$. 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, $F(1,36) = 15.023$, $p = 0.00$, partial $\eta^2 = 0.30$ (large) for insight, and $F(1,36) = 13.43$, $p = 0.001$, partial $\eta^2 = 0.28$ (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.

2.2.3.1. Characteristics of writing-to-learn

In Table 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.

We expected to find differences between the (proportioned) mean frequencies, in favor of the experimental condition.

Table 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* = 0.49 (small effect), revising after finishing an utterance (e.g., *this happens two … this is… no no, this has two reasons*), *d* = 0.75 (medium effect), and rereading own text, *d* = 0.82 (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

TABLE 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	Std. deviation	Mean	Std. deviation	Cohen's *d*
Planning: generating					
Using knowledge about audience	0.00	0.00	0.00	0.00	0.00
Planning: selecting					
Thinking about content selection	0.03	0.04	0.04	0.04	−0.27
Formulating					
Thinking about formulating	0.02	0.02	0.02	0.03	0.00
Revising while formulating	0.03	0.03	0.02	0.01	0.49
Revising after finishing an utterance	0.03	0.04	0.01	0.01	0.75
Monitoring					
Rereading own text	0.08	0.09	0.03	0.03	0.82
Rethinking task approach	0.01	0.01	0.01	0.01	0.00

performed more often by control teacher candidates ($d = -0.27$). The third, thinking about formulating, was equally frequent in both conditions just as 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. Study 2: writing-to-learn in mathematics teacher-training

3.1. Materials and methods

3.1.1. Participants

This study took place in a third-year mathematics teacher-training 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 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 7 presents their characteristics. One mathematics teacher educator was involved in the study. He had 9 years' experience in teacher education and taught both conditions.

TABLE 6 Characteristics of the participants of mathematics teacher education.

n=62	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 7 Characteristics of the selected participants for the think-aloud study in mathematics teacher education.

n=15	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	High school: 3	High school: 4

3.1.2. Design

The design of this study was the same as Study 1. The lessons and writing tasks were about topics from mathematics (rows and limits). Just as in Study 1, the experiment took place in two consecutive academic years and started 1 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.

3.1.3. 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 accessory sums 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. An example of a writing task is to write an explanation of how time and web graphs can be used for visualizing rows of numbers.

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 2.1.3).

3.1.4. Instruments

3.1.4.1. 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 mathematics. For the eight items on topic knowledge, Cronbach's alpha indicated a large heterogeneity (−0.118). The six items on insight were more homogeneous, as indicated by a Cronbach's alpha of 0.56. 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 and we decided to include these tests as covariates in our analysis.

3.1.4.2. 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 0.86, which is considered acceptable.

3.1.4.3. 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 0.15 (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 0.81, which is considered acceptable.

3.1.4.4. 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 they 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.

3.1.5. 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 9 weeks: eight lessons lasting 100 min each and a final examination in the 9th 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 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.

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.

3.1.5.1. 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 selection of (seven) codes as in Study 1 was made as indicating that the process of writing-to-learn is taking place (cognitive and metacognitive processes).

3.1.6. Data analysis

3.1.6.1. 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.

3.1.6.2. Process analysis

The analysis of the process of writing-to-learn was performed as in Study 1 (see Section 2.1.6).

TABLE 8 Lesson structure in control and experimental condition in mathematics teacher education.

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 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

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

3.2. Results

Table 9 shows teacher candidates' mean scores and standard deviations on the pre-tests for prior insight, prior topic knowledge and vocabulary, and the post-test for insight.

TABLE 9 Means, standard deviations for prior insight and topic knowledge, vocabulary knowledge and post-test scores on insight in mathematics teacher education.

N=62	Experimental group (*N* =26)	Control group (*N* =36)
Variables	Mean (SD)	Mean (SD)
Prior insight[a]	4.27 (2.78)	3.85 (2.37)
Prior topic knowledge[b]	3.96 (1.18)	4.17 (1.30)
Pre-test vocabulary[c,d]	14.59 (4.57)	17.11 (6.97)
Post-test insight[e]	28.35 (11.42)	21.22 (11.50)

[a]Theoretical maximum score: 23.
[b]Theoretical maximum score: 8.
[c]Theoretical maximum score: 29.
[d]Four students did not perform the test.
[e]Theoretical maximum score: 49.

3.2.1. Prior knowledge

The differences between students' prior insight in the two conditions were not significant $F(1,60) = 0.414$, $p = 0.522$. The differences between students' prior topic knowledge in the two conditions were also not significant $F(1,60) = 0.406$, $p = 0.527$.

3.2.2. Vocabulary

There were no significant differences between teacher candidates' vocabulary in the two conditions $F(1,56) = 2.27$, $p = 0.137$.

3.2.3. 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 $F(1,56) = 1.829$, $p = 0.182$, $F(1,56) = 0.147$, $p = 0.703$ and $F(1,56) = 3.216$, $p = 0.079$. 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, $F(1,60) = 5.829$, $p = 0.019$, partial $\eta^2 = 0.089$ (medium effect). Experimental students outperformed control students in post-test insight scores.

3.2.4. The process of writing-to-learn

Table 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.

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*), $d = 0.83$ (large effect), thinking about content selection, $d = 0.55$ (medium effect), revising while formulating, $d = 0.38$ (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

TABLE 10 Proportioned means, standard deviations, and effect sizes of indicators of writing-to-learn in experimental (*N*=8) and control group (*N*=7) of mathematics teacher education.

	Experimental		Control		Effect size
Codes	Mean	Std. deviation	Mean	Std. deviation	Cohen's *d*
Planning: generating					
Using knowledge about audience	0.04	0.03	0.02	0.02	0.83
Planning: selecting					
Thinking about content selection	0.03	0.05	0.01	0.01	0.55
Formulating					
Thinking about formulating	0.03	0.02	0.04	0.07	−0.25
Revising while formulating	0.03	0.04	0.01	0.02	0.38
Revising after finishing an utterance	0.02	0.02	0.02	0.02	0.00
Monitoring					
Rereading own text	0.12	0.09	0.04	0.06	1.27
Rethinking task approach	0.01	0.01	0.01	0.02	0.00

formulating was slightly more frequent for the control teacher candidates, $d = -0.25$. The second formulating activity revising after finishing an utterance was performed just as often by both conditions, $d = 0.00$. The monitoring activity rethinking task approach (e.g., *I have to do it in another way…*) was equally frequent in both conditions as well, $d = 0.00$. Thus, in these three cases, the hypothesis was not confirmed.

4. Discussion

4.1. Conclusion

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 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.

TABLE 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.

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,

10.3389/feduc.2023.1094156

leading to more insight and topic knowledge in biology teacher education and insight in mathematics teacher education.

4.2. 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.

4.3. 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 whether 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 (Skemp, 1987; Veel, 1999), 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).

4.4. 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 and 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 and 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 (Bereiter and Scardamalia, 1987; Beal, 1990; 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.

5. 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 Section 6).

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., 2020). 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 et al. (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.

6. Pedagogical implications

Although it appears that writing-to-learn can be applied in many disciplines, it is not much used in education yet (Klein and 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.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Funding

This research received funding from Rotterdam University of Applied Sciences.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/feduc.2023.1094156/full#supplementary-material

References

Akkus, R., and Hand, B. (2011). Examining teachers' struggles as they attempt to implement dialogical interaction as part of promoting mathematical reasoning within their classrooms. *Int. J. Sci. Math. Educ.* 9, 975–998. doi: 10.1007/s10763-010-9266-8

Baaijen, V. M., and Galbraith, D. (2018). Discovery through writing: relationships with writing processes and text quality. *Cogn. Instr.* 36, 199–223. doi: 10.1080/07370008.2018.1456431

Bangert-Drowns, R. L., Hurley, M. M., and Wilkinson, B. (2004). The effects of school-based writing-to-learn interventions on academic achievement: a meta-analysis. *Rev. Educ. Res.* 74, 29–58. doi: 10.3102/00346543074001029

Beal, C. R. (1990). The development of text evaluation and revision skills. *Child Dev.* 61, 247–258. doi: 10.2307/1131063

Bereiter, C., and Scardamalia, M. (1987). *The Psychology of Written Composition*. Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Bos, D., Bouw, J., Brinkman, H., De Bruijn, I., Doekes, W., Van Dijk, B., et al. (2007). Moderne Wiskunde: Vwo-bovenbouw deel B1, B2, B3, D1, D3. 9e editie. Noordhoff.

Boscolo, P., and Carotti, L. (2003). Does writing contribute to improving high school students' approach to literature? *L1-Educ. Stud. Lang. Lit.* 3, 197–224. doi: 10.1023/B:ESLL.0000003685.30925.c4

Boyle, G. J. (1991). Does item homogeneity indicate internal consistency or item redundancy in psychometric scales? *Personal. Individ. Differ.* 12, 291–294. doi: 10.1016/0191-8869(91)90115-R

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences*. Lawrence Erlbaum Associate, Publishers. Academic Press.

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, L. M., and Dunn, L. M. (2005). *Peabody picture vocabulary test-III-NL. L. Schlichting (Dutch version)*. Amsterdam: Harcourt Test Publishers.

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

Emig, J. (1977). Writing as a mode of learning. *Coll. Compos. Commun.* 28, 122–128. doi: 10.2307/356095

Ericsson, K. A., and Simon, H. A. (1993). *Protocol analysis*. London: MIT Press.

Finkenstaedt-Quinn, S. A., Halim, A. S., Chambers, T. G., Moon, A., Goldman, R. S., Ruggles Gere, A., et al. (2017). Investigation of the influence of a writing-to-learn assignment on student understanding of polymer properties. *J. Chem. Educ.* 94, 1610–1617. doi: 10.1021/acs.jchemed.7b00363

Galbraith, D. (1992). Conditions for discovery through writing. *Instr. Sci.* 21, 45–71. doi: 10.1007/BF00119655

Galbraith, D. (1999). "Writing as a knowledge-constituting process" in *Knowing what to write: conceptual processes in text production*. eds. M. Torrance and D. Galbraith (Amsterdam: Amsterdam University Press), 139–159.

Galbraith, D. (2009). Writing as discovery. *Br. J. Educ. Psychol.*, Monograph Series II 6-teaching and learning writing 2, 5–26. doi: 10.1348/978185409x421129

Graham, S., Kiuhara, S. A., and McKay, M. (2020). The effects of writing on learning in science, social studies, and mathematics: a meta-analysis. *Rev. Educ. Res.* 90, 179–226. doi: 10.3102/0034654320914744

Halliday, M. A. K., and Martin, J. R. (1993). *Writing science: literacy and discursive power*. London: The Falmer Press.

Hand, B., Gunel, M., and Ulu, C. (2009). Sequencing embedded multimodal representations in a writing-to-learn approach to the teaching of electricity. *J. Res. Sci. Teach.* 46, 225–247. doi: 10.1002/tea.20282

Hand, B., Wallace, C. W., and Yang, E. (2004). Using a science writing heuristic to enhance learning outcomes from laboratory activities in seventh-grade science: quantitative and qualitative aspects. *Int. J. Sci. Educ.* 26, 131–149. doi: 10.1080/0950069032000070252

Hand, B., Yang, O. E., and Bruxvoort, C. (2007). Using writing-to-learn science strategies to improve year 11 students' understanding of stoichiometry. *Int. J. Sci. Math. Educ.* 5, 125–143. doi: 10.1007/s10763-005-9028-1

Hayes, J. R., and Flower, L. S. (1980). "Identifying the organization of writing processes" in *Cognitive processes in writing*. eds. L. W. Gregg and E. R. Steinberg (Mahwah: Erlbaum), 3–27.

Hohenshell, L. M., and Hand, B. (2006). Writing-to-learn strategies in secondary school cell biology: a mixed method study. *Int. J. Sci. Educ.* 28, 261–289. doi: 10.1080/09500690500336965

Hunter, K., and Tse, H. (2013). Making disciplinary writing and thinking practices an integral part of academic content teaching. *Act. Learn. High. Educ.* 14, 227–239. doi: 10.1177/146978741349803

Kenney, R., Shoffner, M., and Norris, D. (2014). Reflecting on the use of writing to promote mathematical learning: an examination of preservice mathematics teachers' perspectives. *Teach. Educ.* 49, 28–43. doi: 10.1080/08878730.2013.848002

Kieft, M., Rijlaarsdam, G., and van den Bergh, H. (2006). Writing as a learning tool: testing the role of students' writing strategies. *Eur. J. Psychol. Educ.* 21, 17–34. doi: 10.1007/BF03173567

Klein, P. D. (1999). Reopening inquiry into cognitive processes in writing-to-learn. *Educ. Psychol. Rev.* 11, 203–270. doi: 10.1023/A:1021913217147

Klein, P. D. (2004). Constructing scientific explanations through writing. *Instr. Sci.* 32, 191–231. doi: 10.1023/B:TRUC.0000024189.74263.bd

Klein, P. D., and Boscolo, P. (2016). Trends in research on writing as a learning activity. *J. Writ. Res.* 7, 311–350. doi: 10.17239/jowr-2016.07.3.01

Klein, P. D., and Kirkpatrick, L. C. (2010). A framework for content area writing: mediators and moderators. *J. Writ. Res.* 2, 1–46. doi: 10.17239/jowr-2010.02.01.1

Klein, P. D., and Rose, M. A. (2010). Teaching argument and explanation to prepare junior students for writing-to-learn. *Read. Res. Q.* 45, 433–461. doi: 10.1598/RRQ.45.4.4

Klein, P. D., and Samuels, B. (2010). Learning about plate tectonics through argument writing. *Alberta J. Educ. Res.* 56, 196–217. doi: 10.11575/ajer.v56i2.55398

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

Miller, D. M., Scott, C. E., and McTigue, M. (2018). Writing in the secondary level disciplines: a systematic review of context, cognition and content. *Educ. Psychol. Rev.* 30, 83–120. doi: 10.1007/s10648-016-9393-z

Newell, G. E. (1984). Learning from writing in two content areas: a case study/protocol analysis. *Res. Teach. Engl.* 18, 265–287.

Nückles, M., Hübner, S., and Renkl, A. (2009). Enhancing self-regulated learning by writing learning protocols. *Learn. Instr.* 19, 259–271. doi: 10.1016/j.learninstruc.2008.05.002

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

Prain, V. (2006). Learning from writing in secondary science: some theoretical and practical implications. *Int. J. Sci. Educ.* 28, 179–201. doi: 10.1080/09500690500336643

Reece, J. B., Urry, L. A., Cain, M. L., Wasserman, S. A., Minorsky, P. V., Jackson, R. B., et al. (2011). Campbell Biology. San Francisco, Benjamin Cummings/Pearson.

Rose, D. (2008). "Writing as a linguistic mastery: the development of genre-based literacy pedagogy" in *Handbook of writing development*. eds. D. Myhill, D. Beard, M. Nystrand and J. Riley Thousand Oaks: Sage.

Sampson, V., and Phelps Walker, J. (2012). Argument-driven inquiry as a way to help undergraduate students write to learn by learning to write in chemistry. *Int. J. Sci. Educ.* 34, 1443–1485. doi: 10.1080/09500693.2012.667581

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika* 74, 107–120. doi: 10.1007/s11336-008-9101-0

Skemp, R. R. (1987). *The psychology of learning mathematics*. New York and London: Routledge.

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Res. Sci. Educ.* 48, 1273–1296. doi: 10.1007/s11165-016-9602-2

Van Dijk, A., Van Gelderen, A., and Kuiken, F. (2022). Which types of instruction in writing-to-learn lead to insight and topic knowledge in different disciplines? A review of empirical studies. *Rev. Educ.* 10, 1–41. doi: 10.1002/rev3.3359

Van Gelderen, A. (1997). Elementary students' skills in revising: integrating quantitative and qualitative analysis. *Writ. Commun.* 14, 360–397. doi: 10.1177/0741088397014003003

Van Someren, M. W., Barnard, Y. F., and 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 *Pedagogy and the shaping of consciousness: linguistic and social processes*. ed. F. Christie (London: Cassell).

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