Effects of hypertext writing and observational learning on content knowledge acquisition, self-efficacy, and text quality:
Two experimental studies exploring aptitude treatment interactions

Martine A. H. Braaksma°, Gert Rijlaarsdam* & Huub van den Bergh#

°Education Council of the Netherlands | The Netherlands

*Research Institute of Child Development and Education, University of Amsterdam | The Netherlands & University of Antwerp | Belgium

#Utrecht Institute of Linguistics OTS, Utrecht University | The Netherlands

Abstract: In two experimental studies, we examined the effects of types of *written production mode* (hypertext writing versus linear writing, Study 1 and 2) and *learning mode* (performance versus observational learning, Study 2). Participants in Study 1 (Grade 10) were initiating the more formal academic argumentative text, while in Study 2 students (Grade 11) were familiar with the genre. Dependent variables were students' content knowledge, self-efficacy for writing and text quality. For the independent variable *written production mode* both studies did show interaction effects between learning condition and pretest scores. For content knowledge, students with lower prior content knowledge performed best in the hypertext condition; students with higher prior content knowledge in the linear condition. For self-efficacy, linear writing was most effective for students with initial high self-efficacy (Study 2 only). For text quality, students with relatively very strong initial writing skills performed best in the hypertext condition, students with weak initial writing skills in the linear condition (Study 2 only). For the independent variable *learning mode* for the hypertext text learning activity (performing versus observing), almost no differences in effects could be observed: performing the hypertext learning activities or observing these performances did not make a difference, except to students with relatively low initial topic knowledge: students with low prior knowledge performed better in the performing condition. These complex patterns of interactions between learning conditions and pretest variables are discussed.

Keywords: Hypertext writing, observational learning, vicarious learning, intervention studies, content knowledge, self-efficacy for writing, argumentative writing, text quality

1. Study 1: Effects of hypertext writing

1.1 Introduction

It usually takes considerable time and effort for students to build some form of expertise in writing. Similarly, it takes a long while to develop from being an associative writer into a writer who is able to restructure, build and convey knowledge during the writing process (c.f. Bereiter and Scardamalia's knowledge-telling and knowledge-transforming model, 1987).

A major problem for adolescent writers is to write in genres other than lists and narratives consisting of sequenced elements and connected one by one. Text structures like expositions, argumentations and essays require a network of multilinked ideas. Young writers, generating ideas associatively via the knowledge telling strategy, – a mere memory dump – are generally unaware of the hierarchical network of idea units in such texts – and therefore cannot convey this network in their own text (Coirier, Andriessen, & Chanquoy, 1999). Young writers produce ideas as they are generated; they do not reprocess – review and restructure – the generated associative list when producing text. A second problem is the lack of linguistic means – verbal markers, paragraphing, punctuation – to signal the hierarchy of multilinked ideas in the linear text. Readers need these markers to understand the text easily, as it is a reader's task to deconstruct the linear text into an underlying hierarchical information structure (Van Dijk & Kintsch, 1983).

Hypertext writing may contribute to solving both above-mentioned problems. In such a text information is organized as a network in which nodes represent text chunks and links represent relationships between the nodes (Rouet, Levonen, Dillon, & Spiro, 1996). Writing in the hypertext mode requires a well-developed sense of content structure. It relies heavily on ordering, clustering and connecting ideas, irrespective of whether these ideas are generated from memory or from external sources. To arrive at a well-constructed hypertext, the underlying hierarchical information structure needs to be established first. In this sense, a requirement for composing hypertext is to make an in-depth analysis of the hierarchy without thinking in linear formats and to structure the macro elements of a text in a hierarchical rather than linear fashion. Composing hypertext may therefore help students become aware of hierarchical content structures. Paradoxically, this may in turn contribute to students' skills in composing linear texts, for which a solid hierarchical information structure also serves as a good orientation for text production.

The idea that writing hypertexts may enhance learning (and learning to write) is central in Haas and Wickman (2009). They reviewed empirical studies that took the creation of hypertext as its object of analysis and discussed 16 studies that examined writing hypertexts. The focus of these studies differed; some had as object of study to model hypertext writing processes, others focused on the effects of hypertext writing on different learning outcomes (e.g., motivation, design knowledge, content knowledge, creative thinking). Studies that were reviewed in the second category (focus on learning outcomes) took hypertext writing to be a means, not an end; hypertext writing is treated as a complex practice with potentially important pay-offs.

The present study continues to focus on the potential of hypertext writing as a means to enhance learning, in two respects: improving writing skills and improving content knowledge. In general, we hypothesize that hypertext writing is beneficial for the acquisition of content knowledge as well as for writing in term of self-efficacy for writing and linear text quality, especially for learners with a relatively weaker point of departure, be it writing skills or content knowledge. In the next sections, we shall elaborate on these hypotheses for enhancing content knowledge and writing skills.

1.2   Effects on content knowledge

Expectations about the effects of writing on learning are high (Klein, 1999; Klein & Kirkpatrick, 2010). The underlying premise is that the act of writing is a tool for acquiring content knowledge, developing understanding and improving thinking skills. This concept of 'learning through writing' is applied in various disciplines and at various educational levels. The act of writing is seen as a means of transforming the writer's knowledge. The process of knowledge transforming, as Bereiter and Scardamalia (1987) defined it, happens through interaction between content-related knowledge (the topic addressed in the text) and rhetorical knowledge (as reflected by the design of the text and, among other things, by its structure). This dialogical process supports and stimulates text producers to reflect on their knowledge and to restructure and extend it. However, engaging writers in effective knowledge transforming activities is difficult. A meta-analysis of 48 school-based writing-to-learn programs has shown only small average effects on content learning as measured by conventional academic measures (Bangert-Drowns, Hurley, & Wilkinson, 2004).

Theoretically, hypertext writing might be a better candidate than linear writing to stimulate knowledge transforming activities (Lohr, Ross, & Morrison, 1995; Yoshimura, 1998). The hypertext format in itself requires the interaction between rhetorical and content structures and hinders executing the default knowledge-telling strategies displaying in linear text. Producing hypertexts places particular constraints on a document's design in that it affects its non-linear structure, its nodes, its links, its ways of navigation for (different kind of) readers, and so on (Stahl & Bromme, 2004). It is precisely these constraints that create a production process which may induce an effective knowledge-transforming strategy. This production process created by hypertext writing is in line with Klein's (1999) genre hypothesis. This hypothesis claims that when

writers generate content appropriate to each discourse element (e.g., evidence, claims and warrants in the genre of argumentation) and specify the relationships among these elements, they construct corresponding relationships among elements of their own knowledge. This might be especially the case in lengthy, hierarchically structured texts in which each section includes several subordinate propositions (Klein, 1999: 230-231). Because hypertexts, by their very nature, are hierarchically structured with the inclusion of different discourse elements and relationships between these elements, the construction of hypertexts could assist students in writing to learn.

The expectations about the effects of hypertext writing on content knowledge were tested in an experimental study of Bromme and Stahl (2002, 2005). The participants in the experimental hypertext-writing group more often reflected on the semantic structures of the subject area they were dealing with. Moreover, it was shown that structuring the hypertext from two different readers' perspectives supported knowledge gain, particularly in terms of knowledge about the relationships between nodes.

In the studies mentioned above, no indication is made of interactions with learners' aptitudes, attitudes and beliefs. However, there is reason to assume that aptitude treatment interactions are likely to occur. We hypothesize that the effect of hypertext writing depends on students' initial knowledge structures. Thinking prompted by a hypertext format is dialogic, necessary for stimulating the generation of new strings of thought. Hypertext writing may especially support students with low initial knowledge because this prompting by the nodes structure is not hindered by the parallel linearization process (see section 1.3). Two features of hypertexts may stimulate content generation. At a meso text level, writers must create nodes (rhetorical elements) and generate content for these nodes. The nodes function as interaction points: cause-effect, contrast-comparison, argumentation. Interaction stimulates internal dialogue, further thinking, further content generation, not hindered by formulation problems to construct a hierarchy of thoughts in linear texts. Focused associative generation processes within a node are bound to this node, which is a clearer defined element than a paragraph. While creating another node, a hypertext writer may easily go back to a previous node when new content for that node is generated as a side effect. In linear text writing including such a new element is difficult to implement, and therefore the action might be skipped.

1.3   Effects on essay writing

Although theoretical literature (e.g., Bolter, 1998; Lohr et al., 1995; Snyder, 1997; for an overview also see MacArthur, 2006) suggests that hypertext writing might enhance students' writing abilities, no empirical studies have been published about the effects on writing skills after DeWitt (1996) critically analyzed the claims and proposed a research agenda. What is available are papers about lessons in which hypertext writing plays a role (cf. DeWitt, 2001; DeWitt & Strasma, 1999) or non-empirical studies about the implications of hypertext writing (cf. Lohr et al., 1995; Russell, 1998).

We view the effects of hypertext writing on writing from the perspective of "shared" cognitive activities in writing linear texts and hypertexts. What is central in linear writing, as the term suggests, is a linearization process, i.e., transforming subordinated, super-ordinated and coordinated elements of thought into linear text, including cues for readers who must reconstruct this hierarchy from linear text. For hypertext writing, the hierarchicalization process is the key process as it involves breaking down and converting a linearly presented chain of thoughts into elements forming a hierarchical structure. In an earlier study we identified which cognitive activities were involved for secondary school students in hypertext writing when compared with linear writing (Braaksma, Rijlaarsdam, Couzijn, & Van den Bergh, 2002; Braaksma, Rijlaarsdam, & Janssen, 2007). Students performed two linearization tasks and two hierarchicalization tasks under think-aloud conditions. Results showed that planning and analyzing activities during writing contributed to the quality of both hypertexts and linear texts, and that these activities were more often and more readily elicited in hypertext tasks than in linear writing.

Because composing a hypertext apparently offers more opportunities for students to plan and analyze content elements, we hypothesize that stimulating students to write hypertexts might have a positive effect on text quality in linear writing. From the perspective of writing processes, we may see a hypertext as an implicit planning device, supporting semantic clustering and making connections between clusters, in which the connection between clusters is free from language devices. The formulation load is less because formulation is limited to linking elements within a cluster, and does not require the formulation of relations between clusters of genre elements. If the hypertext format supports weaker writers to generate and structure content, and therefore provides them with the experience of more intensive planning, we may expect that they bring this experience into the writing of linear text. Therefore we expect that weaker writers especially will profit from the hypertext experience for their essay writing compared to weaker writers who practice with writing a linear text. Whether these students can or will transfer these learning experiences to linear writing might be an issue.

1.4   The present study

In the present study we investigated the effects of hypertext writing as compared to linear text writing for learning to write and writing to learn. Effects were tested on content knowledge, self-efficacy for writing and linear text quality. In a quasi-experimental pretest – posttest study, we tested the effects of a learning unit in which we varied only one element, the format of the final text to produce: hypertext (experimental condition) versus linear text (comparison condition).

Our research question reads: Is hypertext writing more effective than linear writing for students' 1) content knowledge, 2) self-efficacy for writing and 3) text quality of linear essays, especially for learners who scored initially low on these respective variables? We assume that when students do generate knowledge during writing

lessons for the writing task to be written, and then have to write a hypertext instead of a regular linear text, the hypertext format will stimulate the consolidation of the generated knowledge as we explained in the theoretical framework. The production of a hypertext, requiring less attention to linear formulation, providing more structure on the meso-text level and therefore leaving more control to the learner, will enhance self-efficacy in writing. At the same time, the need to create nodes in a hypertext urges learners to analyze and organize the content in relevant units, and leave them to formulate a smaller amount of text via the linearization process, which provides them with smaller formulation exercises. Therefore we expect that experiencing the structured support of reporting generated content in a hypertext format will facilitate another linear writing task.

Because of the specific content generation and structuring cues a hypertext format induces, while less attention is needed for formulation processes to linearize the content structure, we hypothesize that larger effects on content knowledge, self-efficacy and writing skills will be observed in students with respectively *initial lower content knowledge* (hypothesis 1), *lower self-efficacy* (hypothesis 2) and *lower writing skills* (hypothesis 3) . Therefore interaction effects will be modeled explicitly (cf. Cronbach, 1957, 1975).

2. Method

2.1 Participants

Two experienced male teachers from two schools in the Netherlands took part in the study with five tenth-grade groups (senior general secondary education, age about 16). For each teacher, intact classes were randomly assigned to conditions to avoid confounding teacher effects. Teacher A taught in one of his classes the hypertext writing condition and in two other classes the linear writing condition, teacher B in one of his classes the linear writing condition and in the other the hypertext writing condition. Participating were 61 students in the linear writing condition (75% female students) and 41 students in the hypertext writing condition (66% female students). Students in the tenth grade have written short functional argumentative texts (letters of complaint, for instance) in former grades: the learning unit in this study is their first formal introduction to writing an argumentative text, based on inquiry and documentation.

2.2 Materials

Independent variable: learning unit

We designed and piloted the learning unit in close cooperation with a team of upper secondary school teachers and writing researchers. Materials included a detailed teacher's manual and students' workbooks containing brief theoretical sections, instructions and tasks. We designed the learning unit in two versions: (a) a hypertext

and (b) a linear version. The two versions were identical in text type (argumentative text), theme, number of lessons (five lessons) and instruction time, but differed in just one aspect. We explicitly focused on a minimal but crucial difference between the conditions: only the *text format* of the final text to be produced at the end of the unit differed – linear text or hypertext. Therefore, the first three lessons (70 minutes each) were exactly the same, aiming at content exploration and generation. Then, after 50 minutes in the fourth lesson, the conditions diverged. The students in the hypertext writing condition received brief technical instructions about how to construct hypertexts (20 minutes) and then, in lesson 5, they wrote their argumentative texts in a hypertext format (70 minutes). The students in the linear writing condition, on the other hand, wrote a recipe for a bad mood as a "filler activity" (20 minutes), and in lesson 5 they wrote their argumentative texts in a linear format (70 minutes). In both conditions, students wrote the texts using Microsoft Word. Table 1 presents an overview of the aims and accompanying activities in the learning unit.

Table 1. Overview of the learning unit in the two conditions in Study 1 (each lesson is 70 minutes). LIN: linear writing condition, HYP: hypertext writing condition.

Lesson and phase	Aims and accompanying activities	Condition
1	Aim: acquiring content; activating prior knowledge about charities. Activities: making concept maps about charities, thinking about criteria for charities and writing an extended definition of charities.	LIN and HYP
2a	Aim: acquiring content; concretizing the criteria for charities. Activities: thinking of and inventing a good cause in the neighbourhood of the students' school.	LIN and HYP
2b	Aim: acquiring knowledge about argumentation; experiencing how argumentation works rhetorically. Activities: playing a simulation game. Fundraising to support the proposal for a good cause and convincing other students of the viability of that good cause.	LIN and HYP
3a	Aim: acquiring content about charities. Activities: processing information about charities by reading documentation and taking up a standpoint in a topical issue about charities.	LIN and HYP
3b	Aim: acquiring knowledge about argumentation; structuring argumentation. Activities: selection of arguments to support the adopted standpoint with the help of the documentation and structuring the arguments in an argumentation structure.	LIN and HYP

3c	Aim: acquiring knowledge of presentational aspects of writing. Activities: writing a first version of the introduction to an argumentative text which is rhetorically attractive.	LIN and HYP
4a	Aim: acquiring knowledge of presentational aspects of writing. Activities: practising with presentational aspects of argumentative texts: getting attention from readers, accommodating different kinds of reader perspectives, offering a clearly structured presentation of arguments.	LIN and HYP
4b	Aim: performing a "filler activity", to compensate for the time spent on the technical aspects of hypertext in the HYP condition Activities: writing a recipe for a bad mood.	LIN
4b	Aim: acquiring knowledge of technical aspects of composing **hypertexts**. Activities: practising with technical aspects of composing **hypertexts** in Microsoft Word.	HYP
5	Aim: bringing together all aspects of the lessons series: content, argumentation and presentational aspects. Activities: writing an argumentative text as a contribution to topical issue about charities in **linear** form.	LIN
5	Aim: bringing together all aspects of the lessons series: content, argumentation and presentational aspects. Activities: writing an argumentative text as a contribution to a topical issue about charities in **hypertext** form.	HYP

We chose to focus on argumentative texts based on external sources. This genre requires a critical stance on the topic elements and forces students to create an overt hierarchical network (cf. the hierarchical network of a hypertext) adequate for that genre. In addition, argumentation is one of the major text types included in the national key aims of writing education in upper secondary and is therefore most often tested during final secondary school exams. Formal teaching with respect to writing argumentative texts, however, only starts in grade 10, so the lesson contents were quite new to our participants.

For content exploration, we used the instructional principle of inquiry learning (Hillocks, 1982, 1995). In inquiry learning, students investigate a subject (phenomena) by using basic strategies such as careful observation and representation in language of the phenomena observed, questioning, comparing and contrasting these phenomena with prior knowledge, formulating and testing tentative hypotheses, and finally writing extended definitions of the topic. In lessons 1 and 2, students created concept maps about charities, constructed an extended definition of charities and collaboratively invented a good cause in the neighborhood of the students' school.

The implementation of the lessons was carefully and thoroughly monitored. Lesson observations, log files compiled by the teachers and examinations of student workbooks showed that the lessons were executed as intended. Furthermore, lesson observations and conversations with students and teachers indicated that the learning experiences in both conditions were quite positive. Teachers as well as students liked the topic selected: charities. The teachers appreciated the way in which the lessons and the student activities were organized, the quality of the hypertexts as well as the linear texts, and the pace of the learning unit. The students responded positively to the cooperation with peers, the provided documentation about charities, making the concept maps, collaboratively inventing a good cause, and the practical nature of the learning unit. The students in the hypertext writing condition liked writing the hypertexts; they experienced it as a well-organized way of working. Students in both conditions also made some critical points: they would have liked more time for reading the documentation, and they experienced some repetition (first making a concept map individually and then designing one in small groups).

Dependent variables: testing materials

Content knowledge

To assess the effect of the written production format on the content knowledge involved in the intervention, we used concept maps to measure the organization of content knowledge concerning the topic of the learning unit (charities) at pretest and posttest. Concept maps are useful for investigating students' understandings and planning and for assessing instruction (cf. Good, Novak, & Wandersee, 1990). Students were asked to write as many concepts and relations between these concepts around the (given) central concept charities. We adapted a scoring system developed by Hilbert and Renkl (2008). We scored the number of relations between the concepts which also provided us with information about the number of concepts. Inter-coder reliability (percentage of agreement between two independent coders) for scoring content knowledge was good: 92%.

Self-efficacy for writing

In line with Raedts', et al. (2008) Bandura's (1986, 1997), and Pajares, Hartley, and Valiante's (2001) guidelines, we operationalized students' self-efficacy beliefs as their judgments of competence for the various composition skills connected to the writing task central to this study. We asked students to estimate their writing skill concerning several aspects of argumentative writing that were addressed in the learning unit on a scale from 0 (no chance) to 100 (absolutely certain). The writing self-efficacy scale contained 18 items, for instance: I can search for information in the documentation that I can use for my text; I can write an introduction in which I introduce the issue and in

which I mention my standpoint; I can support my standpoint with strong (as regards content) arguments.

The measure was used at pretest and posttest; reliability scores (Cronbach's alpha) varied from .92 (pretest) to .94 (posttest).

Posttest linear text quality

As a posttest, we measured the quality of a *linear* argumentative text with a new – linear – writing task (topic: broadcasting for youth). Students wrote this text with the help of documentation we provided (mainly newspaper articles) about the issue whether one of the Dutch public channels should be organized especially for young people. Two independent raters holistically assessed the quality of all texts with a mark between 1 and 10 (primary trait). They took into account the requirements that had been presented to the students in the learning unit: persuasiveness, attractiveness and reader awareness. Inter-coder reliability was good: the Spearmann Brown reliability index equaled .92.

Aptitude

As stated before, writing an argumentative essay based on external sources was a new task for the students in this grade. Therefore, implementing a similar pretest task was not a good option; we would have found bottom effects of such a test, and very skewed score distributions. Instead, as a proxy variable, we determined students' aptitude by administering two sections of the verbal intelligence test (DAT, 1984): Cognition of Meaningful Units words and Verbal Analogies (Cronbach's alpha .84). We successfully used this measure as proxy-variable in an earlier study on effects of observational learning on argumentative writing (Braaksma, Rijlaarsdam & Van den Bergh, 2002), albeit that in that study participants were from grade 8.

Correlations between measurements

Appendix A presents the correlations between the measurements. Significant, substantial correlations were only found for corresponding variables between both measurement occasions (content knowledge and self-efficacy[1]). Within the two measurement occasions, no significant correlations between variables were observed. The lack of a correlation between content knowledge and text quality is to be expected, while text quality was measured with a task on another topic than the topic that was used for measuring content knowledge (see section procedures). Unexpectedly, there appeared to be no correlation between verbal aptitude scores and the other measures, even not with text quality (posttest), and no correlation was observed between self-efficacy and text quality. Separate analyses for both conditions showed no difference: no correlations between aptitude and text quality were observed.

2.3   Procedures

We set up a quasi-experimental study with a pre-post-test control group design and two conditions: hypertext writing and linear writing.

Participation in the experiment took place at schools in eight sessions of 70 minutes each. Per week, three sessions were administered (maximum one session per day). During the first session, pretests were administered: the students completed an aptitude test, and a test that measured their self-efficacy for writing.

In sessions 2 to 6, 70 minutes each, students participated in a learning unit about argumentative writing on a current topic: charities, which is a current and topical issue regularly discussed in the news and close to the students' world. The students started in session 2 with making – individually– a concept map about the topic. This concept map was used as a pretest to measure students' initial content knowledge about charities.

In session 7 and 8 posttests were administered. The students completed a test that measured their self-efficacy for writing and produced a concept map about charities to measure their content knowledge (session 7). To assess their writing skill, students wrote in session 8 a (linear) argumentative text about broadcasting for youth, a topic not dealt with in the learning unit.

2.4   Analyses

For all three dependent variables measured in the posttest, we estimated the same type of regression models[2] . In the first model (a so-called null model) it is assumed that there are no differences between conditions, and only an intercept is estimated. In the second model (the corresponding) pretest scores are added to the model (β_1 * Pretest$_i$)[3]. In the third model difference in mean scores between both conditions (linear versus hypertext writing) are allowed. Hence, the differences between the averages of both conditions on the posttest scores are estimated (Δ Hypertext writing; i.e. β_2 * Dummy for Hypertext condition$_i$), while taking differences in pretest scores into account. In the fourth and final model, the relation between pretest and posttest scores is allowed to vary between both conditions (Δ Pretest * Hypertext writing; β_3 * Dummy for Hypertext condition$_i$ * Pretest$_i$) in effect an aptitude treatment interaction between condition and pretest scores is estimated.

The fit of these four nested models can be compared by means of -2log likelihood. The difference in -2loglikelihood of these successive models can be compared by means of the differences in this statistic, which is chi-square distributed with the difference in number of (estimated) parameters as degrees of freedom.

Regression analysis answers the question whether the slopes (the relation between pre and posttest) vary for the different conditions (i.e. β_3). If such an ATI can be shown we know that the slopes in both conditions differ, but not for which range of pretest scores the difference between conditions reaches significance. A special analysis, developed by Johnson and Neyman (1936), is applied to estimate for which pretest

scores the difference between conditions reaches significance ($p < .05$) (compare, Aiken & West, 1991; Pothof, 1964).

3. Results

Table 2 presents the descriptive statistics for all pretests and posttests.

Table 2. Means and standard deviations per condition for pretest and posttest for content knowledge, self-efficacy, verbal aptitude (word knowledge), and text quality in Study 1

| Variable (range) | Linear writing | | | | Hypertext writing | | | |
| | Pretest | | Posttest | | Pretest | | Posttest | |
	M	SD	M	SD	M	SD	M	SD
Content knowledge (0 – ∞)	10.62	3.83	15.03	6.68	8.97	2.20	13.36	4.98
Self-efficacy (0 – 100)	66.62	9.78	72.54	9.15	71.76	8.88	74.45	7.26
Verbal aptitude (0 – 1)	0.59	0.15	NA		0.65	0.15	NA	
Text quality (1 - ∞)	NA		5.28	1.05	NA		5.56	0.95

(NA: not available; M: Mean; SD: Standard deviation).

For both content knowledge and self-efficacy a significant difference between the two conditions on the pretest can be shown (t (53.9) = 2.14; $p = 0.037$; and t (100) = 2.66; $p = 0.009$), with the Linear condition scoring higher on topic knowledge and the Hypertext condition scoring higher on Self-efficacy (Table 2).

Our research questions, however, focus on the difference of the relation between pre- and posttest scores in the conditions. Therefore, we compared the fit of the different regression models. The relevant statistics as well as the comparison of the four models are presented in Table 3.

For content knowledge the fit of the model increases significantly if pretest scores are added to the model ($\chi^2 = 59.3$; $df = 1$; $p < .001$). If condition (Δ Hypertext writing) is added to the model the fit of the model does not improve ($\chi^2 = .1$; $df = 1$; $p = .75$). Adding the interaction between condition and pretest improves the fit of the model significantly ($\chi^2 = 4.7$; $df = 1$; $p = .03$). Hence, the slopes of the regression lines differ between the two conditions.

For self-efficacy only a relation with pretest scores can be shown ($\chi^2 = 27.6$; $df = 1$; $p < .001$), and neither the effect of condition ($\chi^2 = .3$; $df = 1$; $p = .87$), nor the interaction between condition and pretest scores ($\chi^2 = .91$; $df = 1$; $p = .34$) reached significance.

For text quality not even a relation with (verbal) aptitude was significant ($\chi^2 = .89$; $df = 1$; $p = .35$).

Table 3. The fit and comparison of four models for self-efficacy, text quality and content knowledge in Study 1.

Model	-2log likelihood	Comparison			
		Models	χ2	df	p
Self-efficacy					
1. Intercept only	580.84	M1 vs M2			
2. M1 + Pretest[a]	553.21	M1 vs M2	27.63	1	< .01
3. M2 + Condition	553.18	M2 vs M3	.03	1	.86
4. M3 + Condition * Pretest	552.27	M3 vs M4	.91	1	.34
Text quality					
1. Intercept only[a]	291.75				
2. M1 + Pretest	290.86	M1 vs M2	.89	1	.35
3. M2 + Condition	289.33	M2 vs M3	1.53	1	.22
4. M3 + Condition * Pretest	286.28	M3 vs M4	3.05	1	.08
Content knowledge					
1. Intercept only	433.35				
2. M1 + Pretest	383.91	M1 vs M2	49.33	1	<.01
3. M2 + Condition	383.81	M2 vs M3	.11	1	.75
4. M3 + Condition * Pretest[a]	379.08	M3 vs M4	4.73	1	.03

[a]: preferred model

The estimated regression coefficients, which quantify the relation between pretest scores and posttest scores on each variable, are presented in Table 4.

Table 4. Regression coefficients for relation between pre- and posttest scores for hypertext writing and linear writing for content knowledge, self-efficacy and text quality according the best fitting model in Study 1 (β: regression weight; se: standard error).

	Content knowledge		Self-efficacy		Text quality	
	β	(se)	β	(se)	β	(se)
Intercept	14.04[b]	.86	73.57[b]	1.11	5.39[b]	.10
Pretest[a]	3.83	.75	3.90[b]	1.09		
Δ Hypertext writing	-.99	1.35				
Δ Hypertext writing * pretest	-3.45[b]	1.61				
r²	.35		.29		.00	

[a] Standardized; [b]: *p* < .01; [c]: *p* < .02

Content knowledge

In the linear-writing condition the average posttest content knowledge for a student with an average pretest score is estimated as 14.08, which differs significantly from 0 (i.e., $| \beta / se | = 16.33$; $p < 0.01$). In this condition the relation between pretest knowledge and posttest knowledge is significant ($\beta = 3.83$; $se = .75$; $p < .01$); for each point that the pretest score is higher (lower) than the average, the posttest knowledge increases (or decreases) with 3.83. In the hypertext-writing condition the average posttest knowledge of an average student does not differ from that of an average student in the linear-writing condition ($\beta = -0.99$; $p = .23$). However, the relation between pre- and posttest knowledge scores in the hypertext-writing condition is significantly lower ($\beta = -3.45$; $se = 1.61$; $p < .02$) than in the linear-writing condition. The regression coefficient in the hypertext-writing condition is estimated as ($\beta = 3.83 - 3.45 =$) 0.38 which does not differ significantly from 0 ($p = .80$).

As the slopes of both regression lines differ, the relation between pretest knowledge and posttest knowledge differs between both conditions. Therefore, it is assessed for which range of (standardized) pretest scores the difference between both conditions reaches significance. In Figure 1, the (estimated) regression lines for both conditions are plotted. The colored area shows regions at which the difference in knowledge scores
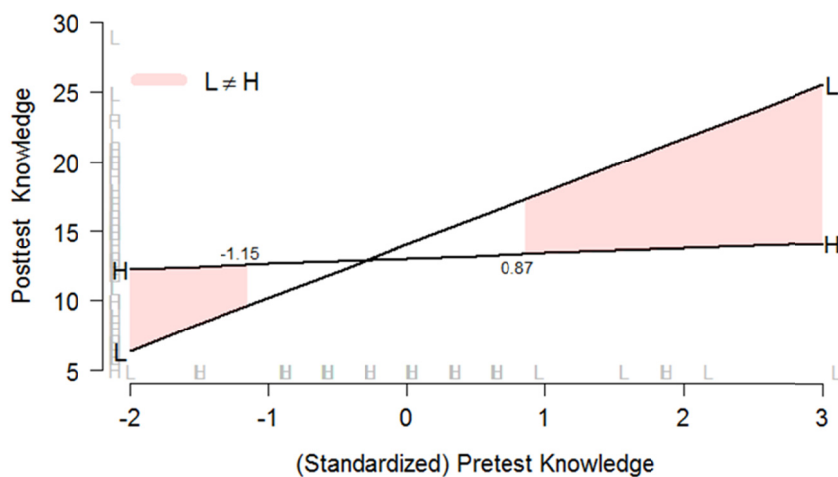

Figure 1. Interaction between pretest content knowledge and condition on posttest content knowledge in Study 1. Estimated regression lines for the relation between pretest and posttest knowledge in the linear-writing (L) and hypertext-writing condition (H).

Note: colored areas indicate $p < .05$ for students with low and high pretest knowledge; grey letters indicate observed pre- and posttest scores in both conditions.

reached significance ($p < .05$). For students with a pretest knowledge score below -1.15, or above .87 the difference in (expected) posttest score differed significantly between conditions. Hence, students with a pretest score of 1.15 *sd* below the pretest mean have a higher posttest score for knowledge in the hypertext-writing condition, and students with a pretest score of 0.87 *sd* above the mean have a higher posttest knowledge score in the linear-writing condition (see Figure 1).

4. Discussion

We examined the effects of hypertext writing as compared to linear text writing on content knowledge, self-efficacy for writing and text quality of linear essays. Results showed no main condition effects on any of the three dependent variables (hypotheses 1, 2, 3), nor did they show the expected aptitude by treatment interaction effects for self-efficacy and text quality (hypotheses 2A and 3A).

What we did observe as expected was an aptitude by treatment interaction for content knowledge (hypothesis 1A). Students with a relatively higher level of initial content knowledge profited more from the learning unit when they had been assigned to the linear writing condition rather than the hypertext writing condition. Students with lower initial levels of content knowledge were better off in the hypertext writing condition than in the linear writing condition. Whether this effect is due to less load on text processing on the meso and macro-level of text as we assumed should be examined more closely in a process study in which writing-to-learn processes demonstrated by students are investigated, preferably with students thinking aloud while composing hypertexts.

Note however, that the format we used to measure topic knowledge might be in favor of the hypertext condition: creating a concept map (measurement) and a hypertext (learning unit) is more alike than creating a linear text. Although this similarity did not result in a main effect of the hypertext condition, it might have supported the students with initial weaker topic knowledge that they create during the learning unit a hierarchical topic model when designing a hypertext.

The absence of expected main effects of production mode on self-efficacy and linear text quality might be due to different factors. For text quality, the lack of effects might be due to the strict instructional design we implemented, with a minimum of learning activities for hypertext writing as a result. We chose to compare the two conditions on the production mode only, to estimate the effect of the text format as purely as possible. Therefore the "only" instruction to the students in the hypertext writing condition concerned technical instruction in the last part of the fourth lesson, which was sufficient to construct their hypertexts, as students indicated and showed with their written hypertexts. It might be possible that transferring the hypertext experience of planning and structuring content to a linear writing task was a bridge too far for students in the case of a first intensive learning unit about a new genre, although one might expect that this would be not the case for the better writers. Some extra

supportive learning activities for hypertext writing and transfer activities could have induced the expected effects. While we cannot compare pretest and posttest writing scores, it is unknown whether both conditions resulted in learning gain regarding the skill to write an argumentative text. If so, we could have concluded that practicing a linear text ('teaching to the test') and practicing a hypertext had equal learning effects.

For self-efficacy such a comparison is possible. We observed an improvement in both conditions of about 0.5-0.6 sd. This effect must be due to the learning unit, while no other writing assignments or lessons were provided during the period the experiment was implemented. When the variation of the production mode between the two conditions does not make a difference in this respect, we tend to ascribe the effect to the quality of the common phase of inquiry learning that provided students with scaffolding activities to acquire content knowledge to write about. We speculate that a possible effect of the production mode, if any, is overshadowed by the effect of the unit as a whole.

Taking into account the above considerations, we planned a study in Grade 11 (pre-university level) in which students had more experience with writing argumentative texts based on sources. This choice would provide us with two advantages. First, we could include a pretest writing task in the research design. Secondly, when students have some experience with argumentative texts, the transfer from a hypertext writing experience to a usual linear writing task might be less far. To raise the chance of transfer in this second study, we added a variant of the experimental hypertext writing condition: an observational learning condition. In this condition students did not perform the learning tasks themselves, but instead observed, compared and contrasted peer models performing these tasks on video. We expected that such a vicarious learning condition would ensure students' *learning* during the production of the hypertext (and thus transferring their new knowledge and skills to linear writing). In the next sections, we will elaborate on these additions and present the set up and results of our second study.

5.   Study 2. Effects of hypertext writing and observation

5.1   Observational learning

Observational learning has been shown to be an effective method for several types of students when it comes to learning to write several types of (linear) texts (Braaksma, Rijlaarsdam, & Van den Bergh, 2002; Couzijn, 1999; Graham & Harris, 1994; Raedts et al., 2008; Rijlaarsdam et al., 2008; Van Steendam, Rijlaarsdam, Sercu, & Van den Bergh, 2010; Van Steendam, Rijlaarsdam, Van den Bergh, & Sercu, 2014; Zimmerman & Kitsantas, 2002). When students learn by observation, they do not execute the intermediate learning-writing tasks themselves but instead observe the learning-to-write processes and the emerging texts of (peer) models. According to Hayes (1996), the long-term memory of writers comprises a set of different kinds of knowledge that ensure

different functions within the writing activity. Knowledge about writing is constructed by means of regulatory or metacognitive strategies such as personal observations, evaluations and reactions (Graham & Harris, 1994). These strategies play a key role in the feedback loop in which one learns from the consequences of one's actions (Zimmerman, 1989, 2000). Via metacognitive strategies, writers gain information that changes what they know and what they do (Graham & Harris, 1994: 209; Oostdam & Rijlaarsdam, 1995). Observational learning supports this use of metacognitive strategies. Students are stimulated to use and address these strategies explicitly, because observing the performance of others involves taking a "natural" step back and thus ensuring a natural type of monitoring, evaluation and reflection upon someone else's task execution processes. Furthermore, cognitive effort shifts from executing writing tasks to learning: students can focus on the learning task, to enlarge their knowledge about writing. Sonnenschein and Whitehurst (1984) convincingly show that observation and evaluation of communicative activities in one mode (listening, speaking) has a strong transfer value to communicative behavior in the other mode.

We claim that observational learning is especially effective in teaching students how to write hypertexts because it is this type of learning that makes it possible to make covert ("mental") processes visible. More specifically: via observational learning, the important planning and analyzing activities during the production of a hypertext can be made visible so that students can learn to perform these activities when they have to engage in writing themselves. Braaksma, Rijlaarsdam, Van den Bergh, and Van Hout-Wolters (2004) showed that observational learning does indeed affect the management of writing processes: compared to students who learned by *performing* writing tasks, students who learned by observation proved to organize their writing processes differently, with more goal-orienting and analyzing activities in the early stages of the writing processes and, overall, more general planning processes.

Vicarious learning is also effective for improving self-efficacy. Schunk and his colleagues (Schunk, 2003; Schunk, Hanson, & Cox, 1987; see also Schunk, 1998: 148) reported effects of (various) models on students' self-efficacy, which in turn influenced learning and achievement.

5.2   The present study

In an experimental pretest – posttest study, with random assignment of students to conditions, we tested the effects of a learning unit aimed at argumentative writing in which we varied the written production task (as we did in Study 1): hypertext writing versus linear writing. Furthermore, for the hypertext writing condition, we created two versions. Based on the hypertext writing condition (Study 1) we created a vicarious learning condition (observational learning), in which learners observed the learning and task activities performed by others instead of performing these activities themselves. As in Study 1 we examined the effects on content knowledge, self-efficacy for writing and linear text quality.

Our research question reads as follows: Which of the three conditions would be the most effective for students' content knowledge, self-efficacy for writing and text quality of linear essays: linear writing, hypertext writing or observational (hypertext) learning? We hypothesize that hypertext writing is more effective than linear writing on content knowledge (hypothesis 1); and that hypertext writing but especially observational learning is more effective than linear writing on self-efficacy for writing (hypothesis 2), and text quality of linear essays (hypothesis 3). For all the variables, we expect an interaction between condition and pretest measures, such that weaker students (content, writing proficiency, self-efficacy) do better in hypertext and observational conditions. Therefore, as in Study 1, we modeled pretest by condition interaction effects for content knowledge (hypothesis 1A), self-efficacy for writing (hypothesis 2A) and text quality (hypothesis 3A).

6. Method

6.1 Participants

The experiment involved 78 seventeen-year-old participants from four eleventh-grade classes (pre-university level) in a Dutch city school. The experiment was part of the school's regular Dutch language and literature lessons. Within classes, students were randomly assigned to conditions: linear writing (N=22, 63% female), hypertext writing (N=26, 62% female) and observational hypertext learning (N=30, 63% female). Under the supervision of a qualified and experienced teacher especially trained for this occasion, the students followed the learning unit delivered in Study 1 in a slightly adapted form. The lessons were provided in the regular schedule, with students of three sub groups (conditions) together in the same classroom.

6.2 Materials

Independent variable: learning unit

Some changes were made in the learning unit on hypertext writing and linear writing which we had delivered in Study 1. These changes are specified in italics in Table 5 (elements that remained unchanged are written in regular font).

The main change compared to Study 1 is the addition of the observational learning condition. Starting in lesson 3, all learning and task activities (except the reading of the documentation in lesson 3, phase 3c) that the participants in the linear writing and hypertext writing conditions had to *perform* were *observed* by the participants in the observational learning condition. With the help of film fragments played on their computer screens, the students in the observational learning group observed peer models performing the activities that students in the other conditions performed. This included, for instance, selecting and ordering arguments from the documentation to

support their standpoint (lesson 4, phase 4a) or creating hyperlinks to their arguments in their final argumentative text (lesson 5).

Table 5. Overview of the learning unit in the three conditions in Study 2 (each lesson is 70 minutes). LIN: linear writing condition, HYP: hypertext writing condition, OBS: observational learning condition

Lesson and phase	Aims and accompanying activities	Condition
1	Aim: acquiring content; activating prior knowledge about charities. Activities: *making a knowledge test about charities*, thinking about criteria for charities and writing an extended definition of charities.	LIN, HYP and OBS
2a	Aim: acquiring content; concretizing the criteria for charities. Activities: selecting the best idea to raise money for a charity (via selling lottery tickets or via a collection).	LIN, HYP and OBS
2b	Aim: acquiring knowledge about argumentation; experiencing how argumentation works rhetorically. Activities: playing a simulation game. Fundraising: developing the idea to raise money for a charity and convincing other students of that idea.	LIN, HYP and OBS
3a	Aim: acquiring knowledge about argumentation: thinking of arguments for a (given) standpoint about charities. Activities: writing two arguments for a given standpoint in **linear** text form.	LIN
3a	Aim: acquiring knowledge about argumentation: thinking of arguments for a (given) standpoint about charities, and acquiring knowledge of technical aspects of composing hypertexts. Activities: writing two arguments for a given standpoint in **hypertext** form.	HYP
3a	Aim: acquiring knowledge about argumentation: thinking of arguments for a (given) standpoint about charities, and acquiring knowledge of technical aspects of composing hypertexts. Activities: **observing** peer models who are writing two arguments for a given standpoint in hypertext form.	OBS
3b	Aim: acquiring knowledge about argumentation and presentational aspects in a text. Activities: commenting on a text with arguments in **linear** text form.	LIN
3b	Aim: acquiring knowledge about argumentation and presentational aspects in a text. Activities: commenting on a text with arguments in **hypertext** form.	HYP

3b	Aim: acquiring knowledge about argumentation and presentational aspects in a text. Activities: **observing** peer models who are commenting on a text with arguments in hypertext form.	OBS
3c	Aim: acquiring content about charities. Activities: processing information about charities by reading documentation and taking up a standpoint in the topical issue about charities (i.e., should commercial lotteries be connected with charities?).	LIN, HYP and OBS
4a	Aim: acquiring knowledge about argumentation; structuring argumentation. Activities: selection of arguments for the standpoint with the help of the documentation and structuring the arguments in an argumentation structure.	LIN and HYP
4a	Aim: acquiring knowledge about argumentation; structuring argumentation. Activities: **observing** peer models who are selecting arguments for the standpoint with the help of the documentation and who are structuring the arguments in an argumentation structure.	OBS
4b	Aim: acquiring knowledge of presentational aspects of writing. Activities: practising with presentational aspects of argumentative texts: getting attention from readers, accommodating different kinds of reader perspectives, offering a clearly structured presentation of arguments.	LIN and HYP
4b	Aim: acquiring knowledge of presentational aspects of writing. Activities: **observing** peer models who are practising with presentational aspects of argumentative texts: getting attention from readers, accommodating different kinds of reader perspectives, offering a clearly structured presentation of arguments.	OBS
4c	Aim: acquiring knowledge of presentational aspects of writing. Activities: writing a first version of the introduction to a **linear** argumentative text which is rhetorically attractive.	LIN
4c	Aim: acquiring knowledge of presentational aspects of writing. Activities: writing a first version of the introduction to a argumentative text in **hypertext** form which is rhetorically attractive.	HYP
4c	Aim: acquiring knowledge of presentational aspects of writing. Activities: **observing** peer models who are writing a first version of the introduction to an argumentative text in hypertext form which is rhetorically attractive.	OBS

5	Aim: bringing together all aspects of the lessons series: content, argumentation and presentational aspects. Activities: writing an argumentative text as a contribution to the topical issue about charities in **linear** form.	LIN
5	Aim: bringing together all aspects of the lessons series: content, argumentation and presentational aspects. Activities: writing an argumentative text as a contribution to the topical issue about charities in **hypertext** form.	HYP
5	Aim: bringing together all aspects of the lessons series: content, argumentation and presentational aspects. Activities: **observing** peer models who are writing an argumentative text as a contribution to the topical issue about charities in hypertext form.	OBS

Note: Changes in the learning unit in comparison with Study 1 are printed in italics

The recordings featuring the models for the observational learning condition were selected from a larger sample. Recordings had been made when we prepared Study 2 with students of age and grade levels similar to those of the participants in Study 2. The models were thinking aloud while performing their tasks. As a result, the observers gained full insight into the models' learning processes. The models were "natural" students at work. Their protocols contained concurrent verbalizations of the activities carried out during task performance, such as planning, analyzing, formulating, transcribing and rereading activities. Some models also offered the observers insight into their monitoring activities, demonstrating monitoring behavior in evaluating and reflecting on their performance.

The participants in the observational learning condition were stimulated to make notes during the observation tasks, as these might help them answer evaluation and reflection questions after observing the models. As previous studies have shown, such questions make sure that students evaluate and reflect on the models' performances in order to enlarge the input for their own learning process (Braaksma, Van den Bergh, Rijlaarsdam, & Couzijn, 2001; Sonnenschein & Whitehurst, 1984). Appendix B presents an example of an observational learning assignment with evaluation and reflection questions.

Next to the addition of the observational learning condition, we made certain other changes in the learning unit originally designed for Study 1 (see text in italics in Table 5). We will explicate one of them here. We moved the technical instructions for making hypertexts from the end of lesson 4 (phase 4b) as planned in Study 1 to the start of lesson 3 (phase 3a) so that students in the hypertext writing condition became attentive to the text type and the skills of hypertext writing at an earlier stage. These technical instructions were now embedded in an assignment with respect to learning how to invent and to write arguments, something which was part of the learning unit.

The implementation of the lessons took place as planned. Observations of the lessons and the student materials showed that the lessons were executed as intended. Moreover, lesson observations and conversations with students as well as the teacher showed that the learning experiences in all three conditions were quite positive, especially the inquiry learning activities in the first two lessons. The students were particularly enthusiastic about the simulation game in lesson 2. Furthermore, similar to Study 1, the students in the hypertext writing condition enjoyed writing the hypertexts. At first, the students in the observational learning condition responded very positively to the observation tasks: they found them inventive and challenging. However, by the end of lesson 5, they indicated that observing peer models (and reflecting on their performances) for more than 200 minutes had become boring.

Dependent variable: testing materials

Following the evaluation of Study 1, certain adaptations in the testing materials were made for use in Study 2.

Content knowledge

In the measurement of content knowledge, we decided to place a greater focus on the topical issue about which they wrote their argumentative texts during the intervention. A more specific task might be a better choice than constructing (broad) concept maps as in Study 1 because we aimed at observing effects of the writing production mode.

The new measurement consisted of six items covering the pros and contras of the topical issue and was identical in the pretest and the posttest. The students had to answer the following two questions: 1) Give three reasons why lotteries as the "Postcode Lottery" are so successful (i.e., three items). 2) Give three point of criticism of lotteries as the "Postcode Lottery" (i.e., three items). A student assistant rated all tests. With each correct answer (i.e., one item), the student earned 2 points, leading to a maximum score of 12 points. Formulation qualities were not taken into account.

Text quality

Argumentative writing tasks for pretest and posttest were taken from a set developed and tested by Van Weijen (2009) and Tillema and colleagues (Tillema, et al., 2013). The topic for the pretest was 'Legalization of soft drugs', for the posttest 'Compulsory automatic donor registration'. Type of assignment, structure, public availability and type of documentation were the same for pretest and posttest; only the writing topic and therefore the content of documentation concerning the topic were different. The writing tasks (topics) were not counterbalanced over pre- and posttests because the generally low correlation between different writing assignments (Coffman, 1966) would have decreased the power substantially.

The students' argumentative essays (pretest and posttest) were rated on global quality with the use of essay scales. Schoonen (2005) demonstrated that holistic ratings (collected with essay scales) have higher generalizability than analytic scores (with scoring guides). Each essay was rated by three raters using one benchmark essay. Topic-specific benchmark essays were available for the pretest and the posttest task. The raters were provided with an explanation of what was "average" about this benchmark essay in terms of criteria that were specified in the instructions for the students, including passages from the benchmark essay. This procedure served to maximize inter-rater reliability. The raters had to award a score to each essay which expressed how much better or worse it was than the benchmark essay (cf. Blok, 1985), which was given the arbitrarily set score of 100. If an essay was awarded a score of 200, for example, this meant that the rater perceived it twice as good as the benchmark essay. If an essay received a score of 50, it meant that the rater perceived it was half as good as the benchmark essay.

For efficiency reasons, we implemented a "design of overlapping rater teams" (Van den Bergh & Eiting, 1989, p. 1). In such a design, the raters do not rate all the texts in the data set. Instead, each rater rates a randomly selected sample of texts. By creating overlap, it is possible to estimate rater reliabilities (see Appendix C). This design was implemented for scoring the pretest and posttest essays. Raters were not aware of learning conditions in this study nor test occasions.

Other measurements

The other tests (verbal aptitude and self-efficacy for writing) used in Study 2 were exactly the same as in Study 1 (see Appendix C for reliability scores).

Correlations between measurements

Appendix D presents the correlations between the measurements. Two out of three correlations for corresponding variables between measurement occasions – self-efficacy and text quality – were significant. In the pretest measurements, a small but significant correlation could be reported between self-efficacy and quality of the written text (r =.33). In the posttest measurements, this correlation had disappeared (overall r =.09), also within each condition separately ($r \leq .28$).

6.3   Procedures

In this study, we opted for an experimental pretest – posttest design, with three conditions and identical pretest-posttests measures for content knowledge, self-efficacy for writing and writing skills.

Participation in the experiment took place at school in seven sessions of 70 minutes each. Per week, three sessions were administered (maximum one session per day). During the first session, pretests were administered: the students completed an aptitude

test and a test that measured their self-efficacy for writing. Furthermore, they wrote an argumentative text (topic: legalization of soft drugs) to assess initial writing skills.

In sessions 2 to 6, students participated in a learning unit (which consisted of five lessons) about argumentative writing (topic: charities). The students started in session 2 with completing (individually) a knowledge test about the topic of the learning unit. This test was used as a pretest to measure students' initial content knowledge about charities.

In session 7 posttests were administered. The students completed a knowledge test about charities and a test that measured their self-efficacy for writing test. Furthermore, they wrote an argumentative text on a new topic (compulsory automatic donor registration) in order to assess their linear writing skills.

6.4   Analyses

In principle the same model as used in Study 1 was applied in Study 2, with one difference: the current model needed two dummies to distinguish between three conditions (linear writing, hypertext writing and observational learning). Hence, in these models the deviation of the intercept for hypertext writing (Δ Hypertext writing) and observational learning (Δ Observational learning) are estimated. Also for the regression of posttest on pretest scores for both the hypertext-writing condition and the observational-learning condition the deviation from the regression coefficient in the linear-writing condition were estimated (i.e., Δ Pretest * Hypertext writing and Δ Pretest * Observational learning).

Table 6. Descriptive statistics for pretest and posttest for content knowledge, self-efficacy and text quality per condition in Study 2

Variable (range)	Linear writing (N = 22)				Hypertext writing (N = 26)				Observational learning (N=30)			
	Pretest		Posttest		Pretest		Posttest		Pretest		Posttest	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Content knowledge (0-9)	2.00	1.17	5.18	2.02	1.69	1.38	4.31	2.06	1.37	0.91	3.63	2.01
Self-efficacy (0-100)	71.63	9.73	74.46	13.39	70.68	11.42	75.50	11.51	74.87	7.26	75.79	9.09
Text quality (0- ∞)	100.19	14.40	101.58	11.55	108.68	15.67	102.14	15.17	96.21	14.17	99.49	17.99

In line with Study 1 it is investigated for which ranges of pretest scores the difference between conditions reaches significance (Johnson & Neyman, 1936; Aiken & West, 1964).

7. Results study 2

Table 6 presents the descriptive statistics for all variables. To test the effect of pretest scores and condition as well as their interaction four models were fitted to the data. The fit of these models per dependent variable, as well as the comparison of the models is presented in Table 7.

Table 7. The fit and comparison of four models for content knowledge, self-efficacy and text quality in Study 2

| Model | -2 log likelihood | Comparison | | | |
		Models	χ2	df	p
		Content knowledge			
1. Intercept only	340.26				
2. M1 + Pretest	371.72	M1 vs M2	18.54	1	<.01
3. 2 + Condition	315.32	M2 vs M3	6.40	2	.04
4. 3 + Condition * Pretest[a]	308.53	M3 vs M4	6.78	2	.03
		Self-efficacy			
1. Intercept only	595.90				
2. M1 + Pretest	553.69	M1 vs M2	42.22	1	<.01
3. 2 + Condition	552.00	M2 vs M3	1.69	2	.43
4. 3 + Condition * Pretest[a]	542.17	M3 vs M4	9.83	2	<.01
		Text quality			
1. Intercept only	629.01				
2. M1 + Pretest	576.25	M1 vs M2	52.76	1	<.01
3. 2 + Condition	573.70	M2 vs M3	2.56	2	.28
4. 3 + Condition * Pretest[a]	564.74	M3 vs M4	8.95	2	.01

[a]: preferred model

For all three dependent variables the interaction between pretest and condition contributes significantly to the fit of the model ($\chi2 > 6.78$; df = 2; $p < .03$). Hence, for all three dependent variables we have to accept Model 4, and assume that the slope of the regression from pre on posttest scores differs between the three conditions.
In Table 8 the estimated regression coefficients are presented for each dependent variable.

Content knowledge

In the linear-writing condition a student with an average score for content knowledge on the pretest is expected to score 5.02. If the pretest score is 1 *sd* above average the

posttest knowledge increases with 0.60 ($p < .01$). In the hypertext-writing condition the average of students with an average content knowledge on the pretest is estimated as (5.02 – 0.83 =) 4.19. The relation between content knowledge on the pretest and content knowledge on the posttest is significantly weaker (β (0.60 – 0.49) = 0.11).

Table 8. Regression coefficients (β) and their standard errors (se) for the relation between pre- and posttest scores in three conditions for content knowledge, self-efficacy and text quality in Study 2.

	Content knowledge		Self-efficacy		Text quality	
	β	se	β	se	B	Se
Linear writing	5.02[c]	.45	72.22[c]	1.78	101.96[c]	2.80
Δ Hypertext writing	-.83	.61	1.15	2.38	-3.05	2.12
Δ Observational learning	-1.13[b]	.57	-.78	2.34	-1.16	1.45
Pretest[a]	.60[c]	.20	11.73[c]	1.75	4.63[c]	1.04
Δ Pretest[a] * Hypertext writing	-.49[b]	.22	-5.34[c]	2.22	3.06[c]	1.16
Δ Pretest[a] * Observational learning	.30	.21	-7.76[c]	2.65	3.90[c]	1.13
r[2]	.38		.50		.32	

[a]: standardized; [b]: $p < .05$; [c]: $p < .01$

Hence, a student with a content knowledge of one standard deviation above the mean on the pretest is likely to have a score of ((5.02 - .83) + (0.60 – 0.49) * 1 =) 4.30. The intercept for 'Observational learning' is significantly lower than in the linear-writing condition. The relation with the pretest scores in this condition does not differ from the slope in the linear writing condition (β = .30; *se* = .21; *p* = 0.16). Here we observe a pure main effect of condition with the linear writing condition scoring higher than the Observing learning condition, regardless the level of prior knowledge. The regression lines for the three conditions are plotted in Figure 2.

Apparent from Figure 2 is that the regression line for the relation between pre and posttest for content knowledge for observational learning is shorter than for linear writing or hypertext writing. This is due to the fact that we didn't observe students with relative high pretest-knowledge score in the observational learning condition (see observed scores in Figure 2). In the observational learning condition the posttest knowledge is always significantly lower than that in the linear-writing condition for students with the same pretest knowledge.

In the hypertext-writing condition the posttest knowledge is higher than in the observational-learning condition for students with a lower pretest content knowledge (i.e., 0.9 *sd* below the mean on the pretest knowledge). For students with a high pretest knowledge no differences between hypertext writing and observational learning were

assessed, because we hardly observed students with high pretest knowledge in the observational-learning condition.
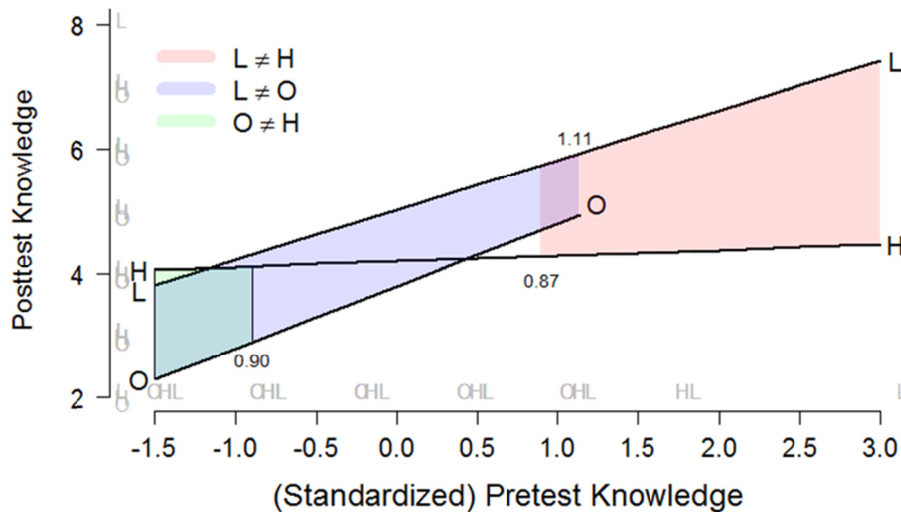

Figure 2. Regression lines for the relation between pre- and posttest content knowledge per condition in Study 2.

Note: L: linear writing; H: Hypertext writing; O: Observational learning. Colored areas indicate significant differences ($p < .05$) between conditions for students with equal pretest scores. Grey letters indicate observed scores in three conditions.

From this Figure it also shows that the difference between linear writing and hypertext writing on posttest content knowledge is only significant for students with relatively higher pretest scores (i.e., ≥ 0.87 *sd* above the mean), and for students with very low pretest scores (< 1.3 *sd* below the mean). In the pretest score interval -1.3 and + 0.9 the scores on posttest knowledge do not differ between the linear and hypertext condition.

Self-efficacy

For self-efficacy no differences between conditions for students with an average self-efficacy score on the pretest ($z_{pretest} = 0$) were found (see Table 8). However, the relation between pretest and posttest self-efficacy clearly differs between conditions. In the linear-writing condition the regression coefficient ($\beta = 11.73$) is larger than in the hypertext-writing condition ($\beta = (11.73 -. 5.34 =) 6.39$) or the observational-learning condition ($\beta = (11.73 - 7.76 =) 3.97$). Further analysis (according Johnson-Neyman, 1936) shows that if the pretest score for self-efficacy is more than 1.06 *sd* above or 0.85 *sd* below the mean the difference between posttest self-efficacy in the linear writing

and hypertext writing condition is significant ($p < .05$); high pretest self-efficacy students have higher posttest self-efficacy scores in the linear writing condition as compared to students with an equal high pretest self-efficacy scores in the hypertext writing condition. On the other hand, students with relatively low self-efficacy scores have lower posttest self-efficacy scores in the linear writing condition as compared to students with equally low pretest self-efficacy in the hypertext writing condition (see shaded area in Figure 3).
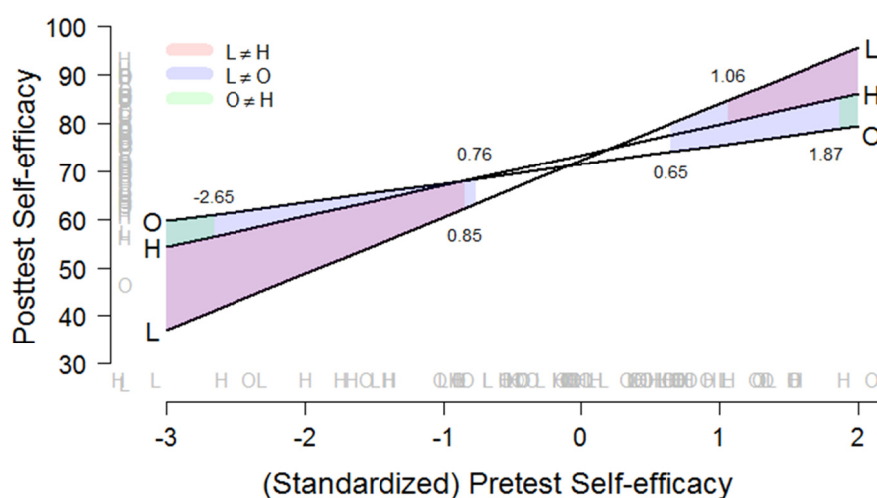

Figure 3. Regression lines for the relation between pre- and posttest self-efficacy per condition in Study 2.

Note. L: Linear writing; H: Hypertext writing; O: Observational learning. Colored areas indicate significant differences (p < .05) between conditions for students with equal pretest scores. Grey letters indicate observed scores in three conditions.

If we compare self-efficacy in the observational-learning condition and linear-writing condition the difference in posttest self-efficacy scores reaches significance if the pretest self-efficacy is more than 0.65 *sd* above or 0.76 *sd* below the mean pretest self-efficacy ($p < .05$; shaded area in Figure 3); students with high pretest self-efficacy scores (more than 0.65 *sd* above the mean) are likely to have higher posttest self-efficacy scores in the linear-writing condition than students with equal self-efficacy scores in the observational-learning condition. Students with low self-efficacy scores (i.e., lower than 0.76 *sd* below the mean) are likely to have higher self-efficacy scores in the observational learning condition than students with equal self-efficacy scores in the linear-writing condition.

A significant difference between posttest self-efficacy scores in the observational-learning and the hypertext-writing condition only shows for a handful of students. Only for students with a pretest score of 1.87 *sd* above of 2.65 *sd* below the mean a significant difference ($p < .05$) in posttest self-efficacy can be shown (see blue-shaded areas in Figure 3). Hence, only for the 0.5% of the students with the lowest self-efficacy and for the highest 3% highest pretest self-efficacy students a difference in posttest self-efficacy can be shown when we compare hypertext writing and observational hypertext writing.

In sum, for self-efficacy the effect of pretest scores depends on the condition; hypertext writing (in both learning modes: practicing as well as observational learning) appears to support weaker students while linear writing supports especially stronger students.

Text quality

For text quality no significant differences between the three conditions in intercept or in regression coefficients could be shown (see Table 8). Hence, the mean text quality in the three conditions does not differ for students with an average text quality on the pretest. The regression coefficient for the influence of pretest on posttest writing is smaller in the linear writing condition than in the other two conditions (see Table 8). As the slopes of the regression lines differ between the linear-writing condition on the one hand and the hypertext-writing and observational-learning condition on the other hand differ, the region for which these differences are significant ($p < .05$) were approximated (Johnson & Neyman, 1936; see Figure 4).
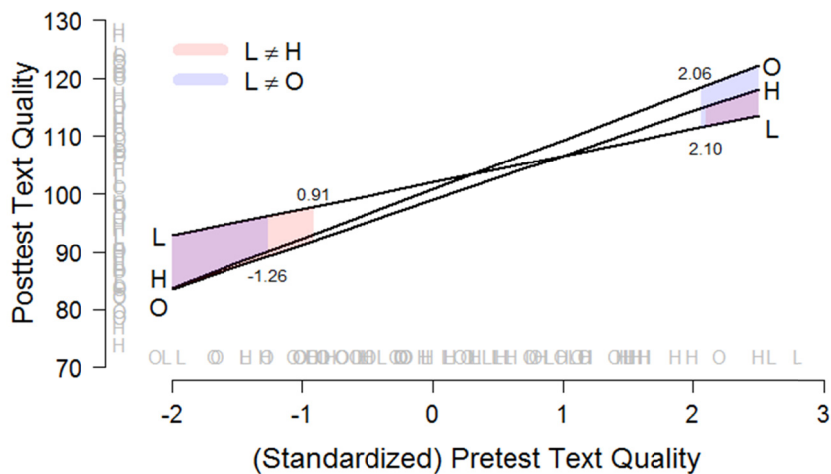

Figure 4. Regression lines for the relation between pre- and posttest text-quality ratings per condition in Study 2.

Note. L: Linear writing; H: Hypertext writing; O: Observational learning. Colored areas indicate significant differences (p < .05) between conditions for students with equal pretest scores. Grey letters indicate observed scores in three conditions..

From Figure 4 it appears that the difference between the linear-writing and the hypertext-writing condition reaches significance if the pretest text quality rating is more than 1.26 sd below the average pretest text quality; the 10% students with the lowest pretest text quality scores are likely to have a higher posttest text quality in the linear writing condition than in the hypertext-writing condition. Students with a very high pretest text quality (i.e., the best 2%) are likely to have a higher text quality in the hypertext-writing condition than in the linear-writing condition.

For students with a relatively low pretest text quality (i.e., 0.91 sd below the mean, or the lowest 18%) the (expected) text quality the linear-writing condition is significantly higher than in the observational learning condition. Also for 2% students with the highest pretest text quality (i.e., 2.06 sd above the mean pretest text quality) the difference between observational-learning condition and the linear writing condition is significant (see Figure 4); a student with a (very) high pretest text quality is likely to produce a text of higher quality in the observational-writing condition than a student with an equal pretest text quality score in the linear-writing condition.

8. SUMMARY OF RESULTS FOUND IN STUDIES 1 AND 2

Table 9 presents a summary of the results found in Study 1 and Study 2. Here, we will clarify this summary; the results will be interpreted and discussed in the following section.

Table 9. Summary of ATI-results found in Study 1 and Study 2.

	Study 1: Initiating argumentative text	Study 2: Familiar with argumentative text
Effect pretest on posttest (over conditions)		
Content knowledge	Yes	Yes
Self-efficacy	Yes	Yes
Text quality	NA	Yes
ATI: interaction effects for various levels of initial scores		
Content knowledge	Yes	Yes
Self-efficacy	No	Yes
Text quality	NA	Yes

Results found in *both* studies indicate:

1. positive effects of the learning unit as a whole on content knowledge and self-efficacy,
2. no main effects of learning condition on content knowledge, self-efficacy and text quality,
3. pretest effect for content knowledge and self-efficacy: students who performed better in the pretest performed better in the corresponding posttest as well in all conditions. This effect was also observed in study 2 for text quality.

In addition, Table 9 shows interaction effects for the effects of pretest on posttest, but these effects *differ* between the studies. In Study 1, the effect of initial content knowledge was larger in the linear writing condition than in the hypertext writing condition; an effect of production mode. In Study 2, we added a condition that varied in learning mode, observational learning, and for none of three variables the results were different from the hypertext writing condition, that operationalized the same production mode, but in another learning mode. Therefore we will only focus on the effects of production mode in Study 2 here.

We found three instances of interaction between learner's initial disposition and condition. First, for content knowledge, linear writing showed to be the best condition for students with relatively high initial scores and hypertext writing was the best condition for students with relatively low content knowledge. For self-efficacy the effect of initial levels proved to be stronger in the linear writing condition than in hypertext writing: for students with an initial low score on self-efficacy, hypertext writing resulted in higher self-efficacy scores, for students with an initial higher score on self-efficacy linear writing resulted in higher scores on self-efficacy. Finally, for text quality, the effect of initial writing skill levels was confirmed for the hypertext writing condition, but not for the linear writing condition: in general, the linear writing condition resulted in the best text quality scores, except for a very small group of initial stronger writers, who scored best in the hypertext writing condition.

9. GENERAL DISCUSSION

9.1 Effects of written production mode: hypertext versus linear writing

In two experiments, we examined the effects of hypertext writing versus linear writing on content knowledge, self-efficacy for writing and linear text quality. In both studies, we found that the effect of learning conditions interacted with learner characteristics.

Content Knowledge. In both studies, we observed an interaction between learning condition and initial knowledge, although the measurement in both studies differ. However, the patterns of the interactions were different.

For students with relatively large initial content knowledge we conclude that the production mode did result in a knowledge effect, for both the measurements of knowledge, free associative knowledge (Study 1), and for the genre specific knowledge

(Study 2). Because of their relatively better knowledge base these students did not seem to need the support of a hypertext writing task to generate and organize ideas (Study 1: associative knowledge) and arguments (Study 2: genre specific knowledge). Instead, hypertext writing may have hindered knowledge building for this group of students. Perhaps, these students, because of their larger and maybe sufficient knowledge base, acquired already enough information during the inquiry lessons, and then the hypertext assignment did not work as an extra dialogic knowledge generator.

For students with relatively low initial content knowledge we conclude that the effect of production mode is not stable over both studies. For this group, writing a hypertext resulted in more content knowledge than writing a linear text (Study 1) while in Study 2 no differential effect was observed for this group. This might be due to the differences in knowledge measurement, if we leave out the option that the difference of the effects is due to grade level (study 1; grade 10, Study 2, Grade 11). In Study 1 content knowledge was measured with concept maps, in Study 2 with a knowledge test (genre specific knowledge). The measurement via concept maps was more closely related to the hypertext condition than to the linear text condition, although this did not result in a main effect, but in an effect for the students with initial weaker levels of topic knowledge. So it might be that in Study 1 indeed the results of the assumed generative power of writing a hypertext (see section 1.2) were observed, but not in Study 2, because of the measurement differences. It might have been that the knowledge about the main arguments and counter arguments (Study 2) were already generated during the first three lessons via the inquiry activities, common to all conditions, and that the writing task did not contribute any more. For Study 1, the content was also acquired during the first three lessons, in both conditions, but the addition of the writing task in two formats made a difference for different groups of students. A more specified theory of the kind of knowledge that is generated by the type of learning activity, in our case, the type of written text production, is necessary.

Self-efficacy. Study 2 revealed a relatively strong interaction between initial self-efficacy and the linear writing condition (see Figure 3). This implies that for students with a high self-efficacy pretest score, linear writing was more effective in terms of self-efficacy than hypertext writing, while for students with a low score on the self-efficacy pretest, hypertext writing was advantageous. The advantage of hypertext writing for students with lower initial self-efficacy scores might be due to the fact that hypertext writing stimulates analyzing and planning activities (Braaksma, Rijlaarsdam, Couzijn, & Van den Bergh, 2002), which may cause an experience of being in control. However, since this effect was not observed in Study 1, we must interpret this result carefully. That the linear writing task resulted for students with a high self-efficacy pretest score in a higher level of self-efficacy than the hypertext writing task, may have to do with their relatively high initial level: doing another linear text, well prepared in the inquiry phase, may have served them well as performance, while doing a hypertext might have been unrelated to their concept of self-efficacy. This explanation would imply that self-

efficacy and writing skills are not strongly connected. This indeed is the case in the present data. In Study 2, in the pretest measurement, a small but significant correlation was observed (r =.33), whereas no relation was observed at the posttest. In Study 1 no relation was observed either.

Text quality. Study 2 revealed that for students with strong initial writing skills, the hypertext writing condition is a better choice than the linear writing condition, but this only holds for the very top ten students. Students with weak initial writing skills profit more from the linear writing condition than the hypertext writing condition. On the basis of earlier research, we expected that writing a hypertext stimulates cognitive activities such as analyzing and planning (Braaksma, Rijlaarsdam, Couzijn, & Van den Bergh, 2002). The frequency of these activities proved to correlate with linear text quality in that study. If we assume that these activities were to some extent triggered in participants, irrespective of their writing abilities, we suppose that only the better writers were able to transfer these experiences to the linear writing task. For students with weak initial writing skills, the hypertext task influenced their self-efficacy positively, but for them this effect did not or did not yet pay out in text quality. The support a hypertext offers in genre specific planning on the macro level of text then may have helped the top 10 students with strong initial writing skills students when they wrote a linear text. They were possibly able to take this experience to a new writing task.

When we combine all effects in a table to guide instructional choices (Table 10), it turns out that the choice for a learning condition depends on the instructional aim and the initial levels of the target skill, attitude or knowledge. For self-efficacy and topic knowledge, weaker students may learn most from hypertext writing, stronger students from linear writing. For text quality, another choice might be more effective: linear writing for most of the students, hypertext writing for the most skilled students. This kind of contradictory choices reminded us of findings of Kieft, Rijlaarsdam and Van den Bergh (2006, 2008) and Kieft, Rijlaarsdam, Van den Bergh and Galbraith (2008). These studies showed that adapting a course design to participants' writing process styles had different effects for different writing styles. When aiming at 'writing to learn', the best option was a course design that accommodates the writing process style: planners must be stimulated to plan, revisers must be stimulated to revise. When aiming that 'learning to write', students with strong preferences (planning, revising) made most progress in writing skill when they participated in a course stimulating pre-planning, while students with weak preferences were best served with a course that stimulated revision. Writing for content learning and writing to improve writing skills seems to require different instructional setting for different groups of learners.

Therefore, setting up studies that study effects on different outcomes (writing skill and one content knowledge) taking in account different groups of learners, will enrich insights in the complex world of learning and writing.

Table 10. Learning Conditions to choose from, provided the initial level of the student. Based on Study 1 and Study 2. LIN: linear writing condition, HYP: hypertext writing condition

Instructional aim	Student's initial level in instructional variable	
	Weaker	Stronger
Content knowledge[study 1]	HYP	LIN
Content knowledge[study 2]	no difference	LIN
Self-efficacy[study 2]	HYP	LIN
Writing skills[study 2]	LIN	HYP

This is the main difference between our studies we reported in this paper and the studies we addressed in our introduction. We focused on effects of hypertext writing on *both* content knowledge *and* writing skills, whereas the other studies focused on one of these aspects only (e.g., Bromme and Stahl (2002) on content knowledge).

To conclude, it seems profitable to provide options for students, which is easy to implement in writing instruction.

9.2 Effects of learning mode: performing versus observing hypertext writing

We examined in the second study the effects of observing hypertext writing versus performing hypertext writing on content knowledge, self-efficacy and text quality. We expected (as addressed in our discussion of Study 1) that observational learning might focus more on the learning aims instead of the aim of completing the writing task, which therefore should enhance the students' self-efficacy and linear text quality. Globally, within the production mode of hypertext writing, observation never produced better scores than practice (see Table 11).

Table 11. Instructional choices (HYP[per]/HYP[obs]) when differentiating between stronger and weaker students in the same course. Based on Study 2. HYP[per]: hypertext writing performing condition; HYP[obs]: hypertext writing observational learning condition.

Instructional aim	Student's initial level in instructional variable	
	Weaker	Stronger
Content knowledge	HYP[per]	no difference
Self-efficacy	no difference	
Writing skills	no difference	

For students with weak initial writing skills, we expected that they would transfer the learning experiences from hypertext to linear text more easily when they would observe instead of constructing a hypertext. The results contradict our assumption that these

students have problems in dealing the double learning agenda. If this were true, stimulating these students to observe, compare and evaluate writing and learning-to-write processes, would improve their linear writing skills. But this does not work out like that. Observation did not improve the transfer from learning to the target task.

For students with strong initial writing skills, we did not observe a differential effect of both hypertext writing conditions either. The benefits of hypertext writing contrasted with linear writing hold for both hypertext writing conditions. That implies that students with strong initial writing skills improved their writing skills more by observing peers involved in constructing a hypertext, instead of writing a linear text. For this group of students the same holds as for the students with weak initial writing skills: observing versus performance is not decisive.

From the responses of students, we learnt that they found observing an interesting activity, but they felt that they got too many of these assignments. This motivational experience might have tempered an initially positive effect. A secondary analysis of the observers' notes and answers to the evaluation and reflection questions in a post hoc study would provide us cues whether initially proficient writers show a different and/or better way of note taking or answering than initially weaker writers (cf. Braaksma et al., 2001).

We assume that initially weaker writers could have learned more from hypertext writing or observational learning if we had been able to give them more support during the lessons. This could have been done, for instance, in a whole class discussion in which some students' hypertexts and different structuring options, navigation possibilities for different types of readers are discussed, and in which the (writing) processes that play a role when writing a hypertext and that should be transferred to linear writing are explicitly addressed. The same holds for observational learning. For weaker writers observational learning might become more effective when one can share and discuss, in groups or plenary in the classroom, the models' (writing) processes, the differences between contrasting model performances, the answers to the evaluation and reflection questions, followed finally by paying explicit attention to the transfer of what students have observed to their own writing. Unfortunately, because of the truly experimental design of the study (three conditions in one class room), these kinds of supporting activities were impossible to realize in our study. Still, for educational practice, we recommend embracing these kinds of activities.

9.3 Suggestions for future research

We recommend designing a process study to examine the writing-to-learn and learning-to-write processes that students perform *during* hypertext writing and linear writing. Such a study could reveal the mechanisms in operation during hypertext writing and examine whether the effects of hypertext writing on linear writing as demonstrated by initially proficient writers might be caused by these students being more strongly engaged in planning and analyzing activities performed while they were writing hypertext, as was found in an earlier study (Braaksma, Rijlaarsdam, Couzijn, & Van den

Bergh, 2002). Moreover, such a process study would allow us to examine whether hypertext writing does indeed require students to engage more actively with the content of the text than linear writing tends to do, as is frequently assumed (cf. Stahl & Bromme, 2004). Furthermore, we recommend setting up an experimental study examining the assumed effects of hypertext writing and observational learning on *writing processes* (measured during the writing of a linear text at posttest). In this way, it can be investigated not only whether hypertext writing and observational learning do indeed affect linear writing processes, but also whether and how a possible different writing approach relates to resulting text quality (cf. Braaksma et al., 2004).

Our last recommendation is to set up and report a series of studies, with partly overlapping designs, instruments and characteristics of participants to increase the level of generalizability and nuanced theory building.

Notes

1. For the Hypertext condition, the correlation between pre- and posttest knowledge was absent.
2. In principle, the data calls for a multi-level analysis of variance because participants are nested within classes and hence separate variance components for the variance between and within classes need to be estimated. However, due to the limited number of classes (and teachers), the variance between classes did not reach significance in any analysis. Therefore, class level was not taken into account in subsequent analyses.
3. For convenience sake pretest scores are standardized.

Acknowledgements

References

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Sage: Newbury Park.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood. Cliffs, NJ: Prentice-Hall.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.

Bangert-Drowns, R. L., Hurley, M, & Wilkinson, B. (2004). The effects of school-based writing-to-learn interventions on academic achievement: A meta-analysis. *Review of Educational Research, 1*, 29-58. doi: 10.3102/00346543074001029

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

Blok, H. (1985). Estimating the reliability, validity, and invalidity of essay ratings. *Journal of Educational Measurement, 22*, 41–52. doi: 10.1111/j.1745-3984.1985.tb01048.x

Bolter, J. D. (1998). Hypertext and the question of visual literacy. In D. Reinking, M.C. McKenna, L. Labbo, & R.D. Kieffer (Eds.), *Handbook of literacy and technology* (pp. 3-13). Mahwah, NJ: Erlbaum.

Braaksma, M. A. H., Rijlaarsdam, G., Couzijn, M., & Van den Bergh, H. (2002). Learning to compose hypertext and linear text: Transfer or interference? In R. Bromme & E. Stahl (Eds.), *Writing hypertext and learning: Conceptual and empirical approaches* (pp.15-38). London: Elsevier Science.

Braaksma, M. A. H., Rijlaarsdam, G., & Janssen, T. (2007). Writing Hypertexts: Proposed effects on writing processes and knowledge acquisition. *L1 – Educational Studies in Language and Literature, 7(4),* 93-122.

Braaksma, M. A. H., Rijlaarsdam, G., & Van den Bergh, H. (2002). Observational learning and the effects of model-observer similarity. *Journal of Educational Psychology, 94,* 405-415. doi: 10.1037/0022-0663.94.2.405Braaksma, M. A. H., Rijlaarsdam, G., Van den Bergh, H., & Van Hout Wolters, B. H. A. M. (2004). Observational learning and its effects on the orchestration of writing processes. *Cognition and Instruction, 22(1),* 1-36. doi: 10.1207/s1532690Xci2201_1

Braaksma, M. A. H., Van den Bergh, H., Rijlaarsdam, G., & Couzijn, M. (2001). Effective learning activities in observation tasks when learning to write and read argumentative texts. *European Journal of Psychology of Education, 1*, 33-48. doi: 10.1007/BF03172993

Bromme, R., & Stahl, E. (2002). Learning by producing hypertext from reader perspectives: Cognitive flexibility theory reconsidered. In R. Bromme & E. Stahl (Eds.), *Writing hypertext and learning: Conceptual and empirical approaches* (pp.39-61). London: Elsevier Science.

Bromme, R., & Stahl, E. (2005). Is hypertext a book or a space? The impact of different introductory metaphors on hypertext construction. *Computers & Education, 44(2),* 115-133. doi: 10.1016/j.compedu.2004.01.003

Coffman, W. E. (1966). On the validity of essay tests of achievement. *Journal of Educational Measurement, 3,* 151–156. doi: 10.1111/j.1745-3984.1966.tb00872.xCoirier, P., Andriessen, J., & Chanquoy, L. (1999). From planning to translating: The specificity of argumentative writing. In J. Andriessen & P. Coirier (Vol. Eds.), *Studies in writing: Vol. 5. The foundations of argumentative text processing* (pp. 1-28). Amsterdam: Amsterdam University Press.

Couzijn, M. (1999). Learning to write by observation of writing and reading processes: Effects on learning and transfer. *Learning and Instruction, 9(2),*109-142. doi: 10.1016/S0959-4752(98)00040-1 1

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist, 12,* 671-684. doi: 10.1037/h0043943

Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist, 30,* 116-127. doi: 10.1037/h0076829

DAT (1984). *Differentiële Aanleg Test.* [Differential Aptitude Test.] Lisse: Zwets & Zeitlinger.

DeWitt, S. L. (1996). The current nature of hypertext research in computers and composition studies: an historical perspective. *Computers and Composition, 13,* 69-84. doi: 10.1016/S8755-4615(96)90036-5

DeWitt, S. L. (2001). *Writing inventions: Identities, technologies, pedagogies.* Albany: State University of New York Press.

DeWitt, S. L., & Strasma, K. (Eds.). (1999). *Contexts, intertexts and hypertexts.* Cresskill: Hampton Press.

Good, R., Novak, J., & Wandersee, J. (Eds.). (1990). Special issue: Perspectives on concept mapping. *Journal of Research in Science Teaching, 27 (10)*. doi: 10.1002/tea. 3660271001

Graham S., & Harris, K. R. (1994). The role and development of self-regulation in the writing process. In D. H. Schunk & B. J. Zimmerman (Eds.), *Self-regulation of learning and performance: Issues and educational applications* (pp. 203-228). Hillsdale, N.J.: Erlbaum.

Haas, C., & Wickman, C. (2009). Hypertext and Writing. In R. Beard, D. Myhill, J. Riley & M. Nystrand (Eds.), *The SAGE handbook of writing development* (pp. 527-544). London: SAGE Publications. doi: 10.4135/9780857021069.n36

Hayes, J. R. (1996). A new framework for understanding cognition and effect in writing. In C.M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences and applications* (pp. 1-27). Hillsdale, N.J.: Erlbaum.

Hilbert, T., & Renkl, A. (2008). Concept mapping as a follow up strategy to learning from texts: what characterizes good and poor mappers? *Instructional Science, 36,* 53-73. doi: 10.1007/s11251-007-9022-9

Hillocks, G. (1982). Inquiry and the composing process: Theory and research. *College English, 44(7), 659-673.* doi: 10.2307/376805

Hillocks, G. (1995). *Teaching writing as reflective practice.* New York: Teachers College.

Johnson, P. O., & Neyman, J. (1936). Test of tests of certain linear hypotheses and their applications to some educational problems. *Statistical Research Memoires, 1,* 57-93.

Kieft, M., Rijlaarsdam, G., & Van den Bergh, H. (2006). Writing as a learning tool: Testing the role of student's writing strategies. *European Journal of Psychology of Education, 21(1),* 17-34. doi: 10.1007/BF03173567

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

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

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

Klein, P. D., & Kirkpatrick, L. C. (2010). A Framework for Content Area Writing: Mediators and Moderators. *Journal of Writing Research, 2(1),* 1-46. doi: 10.17239/jowr-2010.02.01.1

Lohr, L., Ross., & Morrison, G. R. (1995). Using a hypertext environment for teaching process writing: An evaluation study of three student groups. *Educational Technology Research & Development, 43 (2),* 33-51. doi: 10.1007/BF02300471

MacArthur, C. A. (2006). The effects of new technologies on writing and writing processes. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 248-263). New York: The Guilford Press.

Oostdam, R. & G. Rijlaarsdam (1995). *Towards strategic language learning.* Amsterdam: Amsterdam University Press. doi: 10.5117/9789053561560

Pajares, F., Hartley, J., & Valiante, G. (2001). Response format in writing self-efficacy assessment: Greater discrimination increases prediction. *Measurement and Evaluation in Counseling and Development, 33*, 214–221.

Pothof, R. F. (1964). On the Johnson-Neyman technique and some extensions thereof. *Psychometria, 29*, 241-256. doi: 10.1007/BF02289721

Raedts, M., Rijlaarsdam, G., Van Waes, L., Daems, F. (2008). Observational learning through video-based models: Impact on students' accuracy of self-efficacy beliefs, task knowledge and writing performance. In Boscolo, P., Hidi, S. (Eds.), *Studies in Writing, Volume 19, Writing and motivation* (pp. 219-238).Oxford: Elsevier Science and Technology.

Rijlaarsdam, G., Braaksma, M., Couzijn, M., Janssen, T., Raedts, M., Van Steendam, E., Toorenaar, A., & Van den Bergh, H. (2008). Observation of peers in learning to write, Practice and Research. *Journal of Writing Research, 1(1),* 53-83. doi: 10.17239/jowr-2008.01.01.3

Rijlaarsdam, G., Van den Bergh, H., Couzijn, M., Janssen, T., Braaksma, M., Tillema, M., Van Steendam, E., Raedts, M. (2011). Writing. In K.R. Harris, S. Graham, & T. Urdan (Eds.), *APA Educational Psychology Handbook: Application to Learning and Teaching (Volume 3)*. (pp. 189-228). Washington, D.C.: American Psychological Association.

Rouet, J-F., Levonen, J. J., Dillon, A., & Spiro, R. J. (Eds.). (1996). *Hypertext and cognition*. Mahwah: Lawrence Erlbaum.

Russell, G. (1998). Elements and implications of a hypertext pedagogy. *Computers & Education, 31*, 185-193. doi: 10.1016/S0360-1315(98)00029-3

Schoonen, R. (2005). Generalizability of writing scores: an application of structural equation modeling. *Language Testing, 22(1),* 1–30. doi: 10.1191/0265532205lt295oa

Schunk, D.H. (1998). Teaching elementary students to self-regulate practice of mathematical skills with modeling. In D.H. Schunk & B.J. Zimmerman (Eds.), *Self-regulated learning, from teaching to self-reflective practice* (pp. 137-159). New York: The Guilford Press.

Schunk, D.H. (2003). Self-efficacy for reading and writing: Influence of modeling, goal setting, and self-evaluation. *Reading & Writing Quarterly,* 19(2), 159-172. doi: 10.1080/ 10573560308219

Schunk, D.H., Hanson, A.R., & Cox, P.D. (1987). Peer model attributes and children's achievements behaviors. *Journal of Educational Psychology, 79,* 54-61. doi: 10.1037/0022-0663.81.3.431

Snyder, I. (1997). Hypertext. *The electronic labyrinth.* New York: New York University Press. doi: 10.1161/01.HYP.30.1.64

Sonnenschein, S., & Whitehurst, G. J. (1984). Developing referential communication: A hierarchy of skills. *Child Development, 55 (5),* 1936-1945. doi: 10.2307/1129940

Stahl, E., & Bromme, R. (2004). Learning by writing hypertext: A research based design of university courses in writing hypertext. In G. Rijlaarsdam, H. van den Bergh, & M. Couzijn (Vol. Eds.), *Studies in Writing, Volume 14, Effective learning and teaching of writing* (pp. 547-560). Dordrecht: Kluwer Academic Press.

Tillema, M., Van den Bergh, H., Rijlaarsdam, G., & Sanders, T. (2013). Quantifying the quality difference between L1 and L2 essays: A rating procedure with bilingual raters and L1 and L2 benchmark essays. *Language Testing, 30(1),* 1 71-97. doi: 10.1177/0265532212442647

Van den Bergh, H., & Eiting, M. (1989). A method of estimating rater reliability. *Journal of Educational Measurement, 26,* 29–40. doi: 10.1111/j.1745-3984.1989.tb00316.x

Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension.* New York: Academic Press.

Van Steendam, E., Rijlaarsdam, G., Sercu, L., & Van den Bergh, H. (2010). The effect of instruction type and dyadic or individual performance on the quality of higher-order peer feedback in EFL. *Learning and Instruction, 20*, 316-327. doi: 10.1207/s1532690Xci2201_1

Van Steendam, E. Rijlaarsdam, G.C.W., Van den Bergh, H.H., & Sercu, L. (2014). The mediating effect of instruction on pair composition in L2 revision and writing. *Instructional Science*, 1-23. doi: 10.1007/s11251-014-9318-5

Van Weijen, D. (2009). *Writing processes, text quality, and task effects: Empirical studies in first and second language writing.* Utrecht: LOT Dissertation Series.

Yoshimura, F. (1998). On the use of hypertext to enhance awareness of the importance of rhetorical organization. *Journal of Technical Writing and Communication, 28(3),* 227-236. doi: 10.2190/WH77-P2NH-813X-2T5Q

Zimmerman, B. J. (1989). A social cognitive view of self-regulated academic learning. *Journal of Educational Psychology, 81*, 329-339. doi: 10.1037/0022-0663.81.3.329

Zimmerman, B. J. (2000). Attaining self-regulation. A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13-39). San Diego: Academic Press. doi: 10.1016/B978-012109890-2/50031-7

Zimmerman, B. J., & Kitsantas, A. (2002). Acquiring writing revision and self-regulatory skill through observation and emulation. *Journal of Educational Psychology, 94*, 660-667. doi: 10.1037/0022-0663.94.4.660

Appendix A. Correlations between the measurements in Study 1

			Pretest			Posttest		
			2	3		4	5	6
Pretest	1	Aptitude	-.04	.19		.08	.04	.09
	2	Content	-.10			.52**	.10	-.02
	3	Self-efficacy				-.15	.53**	-.03
Posttest	4	Content					.06	.03
	5	Self-efficacy						.03
	6	Text quality						

$^{**}p < .01$

Appendix B. Example of an assignment for students in the observational learning condition (Study 2)

Observation assignment: making (and evaluating) the hyperlink to the arguments

Observe how Eva and Bart are making and evaluating the hyperlink to the arguments. After observing, we shall ask you which student did well and which did less well. You have to explain your answer.

First, open Eva's fragment and watch it.
Then watch Bart's fragment.

Below, you will find room to make notes. Please fill in how many times you watched each fragment.

NOTES

Eva (…. x observed):

Bart (…. x observed):

Make the following assignments after observing Bart and Eva. Indicate your choice (circling 'Eva or Bart' and 'better' or 'less well') in question 1:

1. Compare the students:

Eva / Bart did better / less well

2. Explain your answer

……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………

Appendix C. Reliability for aptitude (verbal intelligence); pretest and posttest self-efficacy, and text quality (Study 2)

Test	Internal consistency (Cronbach's alpha)	Reliability for individual raters	Jury reliability (three raters rating each essay)
Aptitude (14 items)	.64		
Pretest self-efficacy (18 items)	.91		
Posttest self-efficacy (18 items)	.94		
Pretest text quality topic A (global score)		.60 (*SD*.20)	.82 (*SD*.06)
Posttest text quality topic B (global score)		.44 (*SD*.23)	.69 (*SD*.10)

Appendix D. Correlations between the measurements in Study 2.

	Content knowledge	Self-efficacy	Text quality
Content knowledge	.12	-.06	.29[*]
Self-efficacy	-.03	.65[**]	.33[**]
Text quality	.20	.09	.46[**]

[*]$p < .05$; [**]$p < .01$

Note: correlations within the pretest measurements are presented above the diagonal, and correlations within the posttest measurement are presented below the diagonal. At the diagonal, the correlations between corresponding pretest and posttest measures are presented (i.e., self-efficacy, content knowledge and text quality).