Mastering The Art of Academic Writing

van der Loo, Janneke

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

MASTERING THE ART OF ACADEMIC WRITING

COMPARING THE EFFECTIVENESS OF OBSERVATIONAL LEARNING AND LEARNING BY DOING

JANNEKE VAN DER LOO

Mastering The Art of Academic Writing: Comparing the Effectiveness of Observational Learning and Learning by Doing

Janneke van der Loo

Mastering the Art of Academic Writing:
Comparing the Effectiveness of Observational Learning and Learning by Doing

Janneke van der Loo
PhD Thesis
Tilburg University, 2021
TiCC Ph.D. Series No. 76

Cover design by: Eefje Eerden-van der Loo
Print by: ProefschriftMaken

Mastering The Art of Academic Writing: Comparing the Effectiveness of Observational Learning and Learning by Doing

Proefschrift ter verkrijging van de graad van doctor aan Tilburg University

op gezag van de rector magnificus, prof. dr. W.B.H.J. van de Donk, in het

openbaar te verdedigen ten overstaan van een door het college voor

promoties aangewezen commissie in de Aula van de Universiteit op

woensdag 1 december 2021 om 14.00 uur

door

Johanna Maria van der Loo,

geboren te Asten

Promotor: prof. dr. E.J. Krahmer (Tilburg University)

Copromotor: dr. M.A.A. van Amelsvoort (Tilburg University)

Leden promotiecommissie: prof. dr. L. de Wachter (KU Leuven)
prof. dr. G.C.W. Rijlaarsdam (Universiteit van Amsterdam)
prof. L. van Waes (Universiteit Antwerpen)
dr. R. Bouwer (Universiteit Utrecht)
dr. P.A. Bax (Tilburg University)
prof. dr. A.M. Backus (Tilburg University)

Table of Contents

CHAPTER 1

1

Introduction: Learning to Write an Academic Text

Introduction: Learning to Write an Academic Text

Writing plays an indispensable role in education, from primary school to higher education. It is considered a powerful tool for learning: writing makes the learners' ideas visible, and it encourages learners to connect ideas and explore their assumptions (Applebee, 1984). Writing can also positively affect reading and text comprehension (Graham, Gillespie, & McKeown, 2012) which in turn influences learning. In other words, writing can enhance learning (*writing to learn)* which makes transforming learners into competent writers (*learning to write*) one of the main aims of schooling (Graham, 2019).

Learning how to write an effective text typically happens over the course of two decades. Children transform from novice writers applying a knowledge telling strategy in which they write to tell what they know, to adolescent writers applying a knowledge transforming strategy in which they transform knowledge for their own benefit. And, in the case of professional writers, to a knowledge crafting strategy, in which writers craft what they know for the benefit of their reader (Scardamalia & Bereiter, 1987; Kellogg, 2008). However, from studies worldwide, we can conclude that becoming a competent writer is not easy: many learners do not reach a sufficient writing competence, which frustrates attaining educational, professional, and personal goals later on in life (Silliman et al., 2020; Graham, 2019). But why is this the case?

One possible explanation lies in the complexity of writing and the cognitive overload that comes with that complexity. Engaging in a writing task is cognitively very demanding. Writers have to engage, often simultaneously, in different cognitive activities, such as planning, translating, reviewing and monitoring (Flower & Hayes, 1981). This is hard for all writers, but especially for students who have to *learn* how to write. The student becomes so closely involved in the writing process that less cognitive energy is left for learning from that process (Braaksma, 2002). This may lead to frustration for both teachers and their students.

This also seems to be the case in higher education. In higher education, students are required to write on a regular basis, for practice, assessment and for a more general initiation into the writing culture of their scientific discipline. However, despite increased attention for academic writing instruction, university staff in the Netherlands often complains about the (lack of) quality of students' writing: they claim that students are not able to set up a decent reasoning, write incoherent texts without establishing relations within and between paragraphs, do not know when and how to refer to sources, and in some cases even lack basic grammar and spelling skills (Aarts et al., 2019; Bonset, 2010). This raises the question how we can improve academic writing quality, and which instructional methods would be effective for academic writing. In this dissertation we set out to investigate instructional methods for teaching university students to write. More specifically, we focus on (1) the role of observational learning and (2) reflection. We do so by (3) building on and systematically comparing existing methods from studies on observational learning. In the following sections, we will elaborate on the rationale behind these three topics.

1. Learning to write an academic text: observational learning versus learning by doing

Different methods have been proposed to teach students to write an effective text. The most common practice in writing instruction is learning by direct experience, or, learning by doing, in which students actively engage in executing writing tasks. In writing research, learning by doing has often been compared to observational learning (see e.g., Raedts et al., 2007; Braaksma et al., 2002; Braaksma et al., 2004). Observational learning occurs when people learn from observing the behavior of others (Bandura, 1997), and has shown positive effects on learning different types of tasks, such as motor skills (e.g., dart throwing), highly structured cognitive tasks (e.g., math), self-regulatory skills, but also ill-structured cognitive tasks, such as writing (Van Gog & Rummel, 2010).

In observational learning, models, (e.g., a teacher, an expert or a peer) demonstrate a problem-solving or performance procedure to a learner. In the case of cognitive tasks, such as writing, the models not only display the strategies and sequences that can lead to a successful performance, but also verbalize their thoughts on the task performance ("I am going to re-read my sources") or on affective factors ("I think this is really difficult") (Schunk, 2003). Observational learning gives students not only the opportunity to observe, but also reflect on and emulate models, and is an instructional method that effective instructors use to teach planning, revising and editing strategies, and self-regulation strategies (Graham & Perin, 2007).

The key difference between learning to write by observation and learning by doing, is that in observational learning the learner does not engage in the cognitive complex and demanding task of writing itself, but instead can focus and reflect on the strategies and the task performance displayed by the model. It could therefore be argued that the cognitive effort is shifted from performing the task to actually learning from the task.

Observational learning has been studied quite extensively in the writing domain with different types of tasks and different types of audiences (see e.g., Braaksma et al., 2002; Zimmerman & Kitsantas, 2002; Rijlaarsdam et al., 2008; Van Steendam et al., 2010; Groenendijk et al., 2013; Fidalgo et al., 2015). However, most of these studies focus on younger students from grade 1-12 (Graham et al., 2016): the role of observational learning in learning to write complex tasks by university students is understudied. In writing at university level, the complexity of the writing task (in length, structure and content) increases and students have to deal with new genres, such as a scientific research report. Only a handful of studies by Raedts and colleagues have explicitly studied observational learning to write with university students (e.g., Raedts, Rijlaarsdam, Van Waes & Daems, 2007; Raedts, Daems, Van Waes, & Rijlaarsdam, 2009). These studies investigated the effects of observational learning with students on task knowledge and text quality compared to learning by doing. The studies showed that the texts written by students who learned by observation were of higher quality: they linked the source material more often, and wrote better organized texts compared to the students in the learning by doing condition. In addition, students who learned by observation were also better calibrated for self-efficacy than the students who learned by doing. Observational learning also led to significantly more extensive knowledge on effective writing strategies.

These findings are a first indication of the possible effectiveness of observational learning as an instructional method for university students. But since the evidence is still rather scarce, we will study this further in this dissertation. This leads to the first research question of this dissertation:

Research Question 1: What is the effect of observational learning on text quality and self-efficacy in the context of academic writing compared to learning by doing?

In addition, a closer inspection of the methods from previous studies reveals an important difference between observational learning and learning by doing, namely reflection. In most studies on observational learning, students who learn by observation are actively prompted to reflect on the models' behavior, while reflection is mostly absent in learning by doing. This raises the question what role reflection plays in the possible effectiveness of observational learning.

2. Reflection in (observational) learning to write

Reflection engages individuals in exploring their experiences in order to lead to new understandings and appreciations: by reflecting learners "recapture their experience, think about it, mull it over, and evaluate it" (Boud et al., 2005), which helps them to consciously think about what kind of actions they will or will not engage in in the future, making it a powerful tool for learning. Reflection is an inherent part of observational learning (Graham & Perin, 2007): for example, in the studies by Raedts and colleagues, students actively reflected on the models' behavior by answering questions in which they had to evaluate and elaborate on the models' actions. This reflection draws the learners' attention towards relevant cues, and arguably supports them in developing a mental representation of the desired performance and in developing performance strategies which might lead to a more successful performance (Ste-Marie et al., 2012). Since reflection is mostly absent in learning by doing in studies comparing observational learning and learning by doing, it is conceivable that differences between these two instructional methods are closely related to presence or absence of reflection, even though this has not been studied before.

Several studies report positive effects of reflective activities in general (i.e. not only combined with observational learning) during learning to write on writing performance, for example in curricula directed at promoting strategies for self-regulated writing (e.g., MacArthur et al., 2015; Koster & Bouwer, 2016; Raedts et al., 2017). In these studies, reflection is of necessity somewhat different from reflection in observational learning studies. In the observational learning studies, reflection is aimed at the models' actions. In other studies, reflection is however directed at the learner's own behavior and therefore also commonly referred to as self-reflection. For example, writers might monitor their progress toward their writing goals, in order to gain a more effective control of their performance strategies (*self-observation*), evaluate their performance by comparing it to standard or goal (*self-judgment*), or respond to their performance outcomes (*self-reaction*) (Zimmerman, 1989; Chang, 2007). A number of studies have shown that self-regulated learning curricula, including one or more of these reflective activities, are positively related to learning to writing competence with different types of audiences (e.g., Zimmerman & Kitsantas, 2002, 1999; Graham et al., 2005; MacArthur et al., 2015).

These studies imply that reflection, either in combination with observational learning or not, may positively influence writing performance, and therefore reflection might complicate the comparison between observational learning and learning by doing in the previous studies by Raedts et al. Could the differences between observational learning and learning by doing (partly) be attributed to reflection? In this disseration, we therefore investigate the role of reflection in learning to write by systematically comparing observational learning and learning by doing, both with and without reflection. This leads to our second research question:

Research Question 2: How does reflection affect observational learning and learning by doing in academic writing?

3. Evidence-informed practices for learning to write for university students

Finally, a more general question is how and when we can determine that an instructional method is actually effective. We set out this dissertation in order to investigate effective ways to teach students to write complex texts, in order to hopefully contribute to a growing body of evidence-based, or evidence-informed, instructional writing practices in higher education. According to Graham et al. (2016) evidence-based writing practices emerging from research are a more reliable source of information, compared for example to the professional opinion and expertise of the writing instructor, because of the generalizability and the reliability of these findings. However, there has been some discussion in how to establish whether an instructional method is actually evidence-based (see e.g., Carter, 2010; Hjørland, 2011). There does seem to be consensus on the fact that the effectiveness of an educational practice should at least be established in multiple studies (Hjørland, 2011). Therefore, meta-analyses are considered a valuable approach.

In the writing domain, several meta-analyses originate from Graham, Harris, Perin and colleagues (see e.g., Graham & Perin, 2007; Graham, Harris, & Chambers, 2016). Cognitive modeling, i.e., observational learning, has indeed been identified as (part of) evidence-based practices in their meta-analyses. Even though this sounds promising for the effectiveness of observational learning (and reflection) in learning to write at university level: the studies included in the meta-analyses by Graham and colleagues were mostly based on samples of younger students (grade 1-12), which means the findings cannot simply be generalized to university students. In addition, meta-analysis as a research method has its limitations. Based on an array of previous studies, Makel and Plukel (2014) identify several of these limitations, such as publication bias towards positive findings and related to that, the 'file drawer problem' which relies on the assumption that statistically non-significant results are less likely to be published and less likely to be included in meta-analyses. They therefore argue that to accumulate understanding in educational practices, researchers should more frequently conduct replication studies.

Replication is the purposeful repetition of previous research to corroborate or disconfirm previous findings and allows us to evaluate the rigor and robustness of scientific findings (Zwaan et al., 2018). Two types of replication are often distinguished: direct replications and conceptual replications. Zwaan et al. (2018) define a direct replication as "a study that attempts to recreate the critical elements (e.g., samples, procedures, and measures) of an original study where those elements are understood according to a theoretical commitment based on the current understanding of the phenomenon under study, reflecting current beliefs about what is needed to produce a finding" (see also Nosek & Errington, 2017). Direct replications are often contrasted with conceptual replications in which researchers "attempt to replicate an important original finding while acknowledging differences in background factors (e.g., participants, details of procedures) compared with the original article" (Lynch et al., 2015). This means that in conceptual replication the same construct-to-construct relations as the original article are studied, even though they might be operationalized in a different manner (Lynch et al., 2015).

In many research domains, including the educational sciences but also the writing domain, the number of both direct replications and conceptual replication studies is scarce (Makel & Plucker, 2014; Plucker & Makel, 2021). Recent developments have, however, increased attention for the need for replication studies, such as the efforts of the Open Science Collaboration (2015). In order to find evidence for the rate and predictors of reproducibility, they replicated 100 results from top-tier journals in psychology. In their attempts, they found that only 39% of the replication efforts yielded significant findings and 32% of the original findings were no longer significant when combined with the new data, which raises concerns on the generalizability of theoretical claims based on these studies.

This is not unique to the field of psychology. Within the educational sciences, only a mere 0.13% are direct replication studies, and only about half of these replications were able to successfully replicate the original findings (Makel & Plucker, 2014). This is concerning since with these numbers it is very likely that at least, similar to psychology, a number of educational claims are based on false positives (Makel et al., 2019). This highlights the importance of replication in education research. As Makel and Plucker (2014) state: "If education research is to be relied upon to develop sound policy and practice, then conducting replications on important findings is essential to moving toward a more reliable and trustworthy understanding of educational environments."

In this dissertation, we therefore chose to conceptually replicate previous studies, in order to establish the effectiveness of observational learning and reflection in learning to write at university level. We attempted to follow the design, materials and procedures as closely as possible. In the next section we will foreshadow the four studies of this dissertation.

Reading Guide for the Chapters

In the four studies of this dissertation we focus on the comparison between learning by doing and observational learning in relation to academic writing performance. In addition, we investigate the role of reflection in both learning by doing and observational learning. This leads to the following research questions: (1) What is the effect of observational learning on text quality and self-efficacy in the context of academic writing compared to learning by doing? (2) How does reflection affect observational learning and learning by doing in academic writing? To answer these questions, we set up four studies which are all conceptual replications. We chose conceptual replications in order to evaluate the theoretical concepts behind the previous and our own findings, providing us with information about the generalizability of the findings (Zwaan et al., 2018).

The different chapters have previously been published as separate journal articles, which means there is some overlap between the theoretical frameworks and methods of, in particular, Chapter 2 and 3, and Chapter 4 and 5.

In *Chapter 2* we report the findings of a study in which we compare observational learning with learning by doing in university synthesis writing. We wanted to test, in an actual first-year bachelor course on academic writing, whether observational learning leads to higher quality texts and whether it affects self-efficacy. The study is a conceptual replication of Raedts et al. (2007) and closely follows the procedures of this study.

From the first study, we conclude that in the comparison between observational learning and learning by doing, there might a confounding factor. In observational learning of a cognitive skill, learners are often asked to reflect on the models' behavior explicitly and actively, while it is mostly absent in the learning by doing activities. Since reflective activities in general arguably enhance learning, it is possible that reflection mitigates the effects of observational learning. In *Chapter 3* we therefore report on a study in which we systematically compare observational learning and learning by doing, both with and without reflection. This study is a conceptual replication of both Raedts et al. (2007) and our study from Chapter 2.

The two single subject design studies from Chapter 2 and 3 were conducted in a 'real life' writing course. This supports the ecological validity of the studies, but it complicates controlling for other factors that might be of influence. In *Chapter 4* we therefore attempt to study observational learning and reflection in a more controlled environment (a lab study) and with a less complex writing task (sentence combining), by conceptually replicating a study by Zimmerman and Kitsantas (2002).

In Chapters 2, 3 and 4 we focus on cognitive writing tasks. In *Chapter 5* we investigate the effects of observational learning and reflection in a completely different domain, motor skill acquisition, by conceptually replicating a study by Zimmerman and colleagues on acquiring dart throwing skills (Kitsantas, Zimmerman, & Cleary, 2000). This allows us to test to what extent our findings from a cognitive domain transpose to the motor skill domain.

Finally, in *Chapter 6*, we will summarize the main findings from each study. Furthermore, we will discuss methodological issues that arose from our results, provide suggestions for further research, and propose some implications for education.

CHAPTER 2

2

Learning How to Write an Academic Text: The Effect of Instructional Method and Writing Preference on Academic Writing Performance

This chapter is based on:
Van der Loo, J., Krahmer, E., & Van Amelsvoort, M. (2018). Learning How to Write an Academic Text: The Effect of Instructional Method and Writing Preference on Academic Writing Performance. *Journal of Writing Research*, 9(3). https://doi.org/10.17239/jowr-2018.09.03.01

Abstract

In this study we investigated which instructional method is suitable for university students to learn how to write an academic text. We have compared observational learning with learning by doing, and we have explored the effects of writing preference (planning versus revising) on academic writing performance. In an experiment 145 undergraduate students were assigned to either an observational learning or learning by doing condition. In observational learning participants learned by observing a weak and strong models' writing processes. In learning by doing they learned by performing writing tasks. Prior to the sessions participants were labeled as either planners or revisers based on a writing style questionnaire. The effects of the sessions were analyzed with a 2x2 between-subjects design with instructional method (observational learning, learning by doing) and writing preference (plan, revise) as factors. To measure academic writing performance the participants wrote an introduction to an empirical research paper. We found no main effects for instructional method and writing preference. Simple effect analyses did reveal that revisers benefitted somewhat more from observational learning than planners. Planners performed equally well in observational learning and learning by doing. However, planners who learned by doing did seem to outperform revisers who learned by doing. Our study suggests that observational learning presents interesting opportunities for academic writing courses. However, more research on the interplay between writing strategy and instructional method is called for.

Introduction

Academic staff often express concern about the inability of students to review, integrate and synthesize scientific literature (e.g., Granello, 2001). Over the years there have been several approaches to instructional methods for improving these academic writing skills. In the current study we compare two of these methods, observational learning and learning by doing, for learning how to write an academic text.

Writing a coherent and effective text is a complex and demanding task (Kellogg, 2008). Why is writing a text so difficult, and hard to learn? One possible explanation is cognitive overload (Braaksma, 2002). During writing there is a complex interplay of four main cognitive activities, namely planning (generating ideas, organizing information and setting goals), translating (putting ideas into language), reviewing (evaluating and revising text) and monitoring (deciding when to move from process to process). Writers have to attend to all these components, often simultaneously (Flower & Hayes, 1981). They have to carry out different processes and pay attention to many textual characteristics simultaneously and therefore may lose track of their own thoughts during the process (Braaksma, 2002). This applies to all writers, but especially to those who are learning to write. When learning to write, the learner becomes so closely involved in the writing process that hardly any cognitive energy is left for learning from that process. A method that allows for a distinction between writing and learning to write, and explicitly allows for reflective activities and a direct link between writing processes and the resulting writing product is observational learning (Braaksma, 2002).

Observational learning occurs when people learn from observing the behavior of others (Bandura, 1997). In his Social Learning Theory, Bandura (1977) describes four sub-processes that lead from observation of modeled events to a matching pattern of behavior. Firstly, observers have to be attentive to the modeled behavior. Secondly, observers must be able to remember specific characteristics of the behavior in order to imitate it. Thirdly, in the reproduction sub-process observers organize and rehearse the modeled behavior symbolically and then enact it overtly. Finally, the decision to reproduce (or refrain from reproducing) an observed behavior is dependent on the motivations and expectations of the observer. This four-step model was also identified in the writing domain in a study by Braaksma, Van den Bergh, Rijlaarsdam and Couzijn (2001) on observation tasks that were effective when students were learning to read and write argumentative texts for the first time.

In learning to write, an important difference between observational learning and other approaches, such as most of the training available at universities (learning by doing), is the lack of actual writing. In observational learning learners do not write themselves but observe and reflect on the writing processes of a model and the emergence of the resulting texts, which demonstrates the complexity of the writing process. By observing, the cognitive effort is arguably shifted from executing writing tasks to learning (Couzijn, 1999; Rijlaarsdam & Couzijn, 2000; Braaksma, 2002).

Various studies have shown that observational learning is effective in the domain of writing (among various other school subjects) with learners of various ages (e.g., Zimmerman & Kitsantas, 2002; Braaksma, Rijlaarsdam, & Van den Bergh, 2002; Rijlaarsdam et al., 2008; Raedts, 2008*).* For example, Couzijn and Rijlaarsdam (2004) compared observational learning with learning by doing in an experimental course for 9th grade students on composing an argumentative text. Participants who learned by doing were presented with theoretical knowledge on writing argumentative texts. Based on the theory, they executed several writing assignments. In the observational learning condition participants were presented with the same theory. However, instead of performing the assignments themselves, the participants observed two peer models executing the tasks while thinking aloud. Before observing the participants were instructed to aim their attention to evaluating the models' performance so they could check whether the models applied the theory correctly. After the observation, the participants had to determine if one model did worse than the other and explain what exactly made this performance less successful. In this way the participants were forced to designate 'strong' and 'weaker' models. Couzijn and Rijlaarsdam (2004) found that participants observing peer models outperformed students who learned by doing on writing an argumentative text.

Braaksma, Rijlaarsdam, Van den Bergh, & Van Hout-Wolters (2004) suggest that the effectiveness of observational learning in learning to write is a result from the observers' strong engagement in metacognitive activities. Observers internalize, apply, and develop criteria for effective writing, by observing the models' writing, identifying, and conceptualizing the writing strategies, evaluating the performance of the models and reflecting explicitly on the observed performances.

Would observational learning also be a suitable method for university students to learn how to write an academic text? Only a small handful of studies, all by Raedts and colleagues, have asked this question (e.g., Raedts, Rijlaarsdam, Van Waes, & Daems, 2007). These studies investigated the effects of observational learning with undergraduate students on task knowledge and text quality by comparing observational learning with learning by doing. During the first session all students were presented with theoretical knowledge. Afterwards, students who learned by doing performed writing exercises while students who learned by observation watched videos in which a weak and strong model performed these exercises. After observing the students were asked to identify the stronger model and they had to write down the strategies used by the models. The studies showed that, contrary to the expectations, students in the observational learning condition did not have more detailed knowledge of what a good literature review should look like. However, their knowledge of effective writing strategies was significantly more extensive. More specifically, there were effects of instruction for strategies concerning information gathering and planning of the text, but no effects for strategies concerning text production and revision. With regards to text quality, the studies showed that students in the observational learning condition outperformed those in the learning by doing condition. The students in the observational learning condition linked the source material more often and wrote better organized literature reviews compared to the students in the learning by doing condition (Raedts et al., 2007).

The work of Raedts and colleagues offers a first indication that observational learning may be a useful strategy for students to learn how to write academic texts. However, various questions remain open, such as the influence of specific learner characteristics. Students for example may differ in writing preferences. Galbraith and Torrance (2004) distinguish between two different approaches to writing common in research: a planning approach and an interactive approach. In the planning approach writers concentrate on working out what they want to say before they actually start producing full text. After establishing what they want to say, they work on expressing the meaning they want to convey as effectively as possible. This approach can include activities such as creating a thinking scheme and planning the text by writing an outline (Kieft, Rijlaarsdam, Galbraith, & Van den Bergh, 2007). In the interactive approach writers work out what they want to say while writing and the content evolves over a series of drafts (Galbraith & Torrance, 2004). According to Elbow (1998) the interactive approach allows writers to look at writing as “an organic, developmental process in which you start writing at the very beginning.” This approach can include several prewriting activities such as free-writing (Elbow, 1998) and involves writing multiple drafts (Murray, 1980). Galbraith and Torrance (2004) indicate that a planning approach may be equally effective as an interactive approach. They suggest that, in general, in education it may be necessary to accommodate different cognitive styles, since there seem to be individual differences in preferences for different strategies. Galbraith, Torrance, and Hallam (2006) found that students benefitted most from writing instruction opposite to their preferred strategy. The writing instruction helped to supplement and foster those strategies which students by themselves did not prefer.

Kieft, Rijlaarsdam, and Van den Bergh (2008) also studied the effect of adapting a writing course to students’ writing strategies. They refer to the interactive approach as a revising strategy. In their study, they found an interaction between writing preferences and writing instruction. Students with a strong writing preference, either planning or revising, learned more from a writing course that was adapted to their writing preference. Rijlaarsdam et al. (2008) have suggested that adaption of observational learning tasks to students’ writing preference therefore may be useful. Students with a planning preference might benefit from observational tasks providing feedback on planning problems while students with a revision preference might benefit more from observations of writing students coping with revision problems. However, based on Galbraith et al. (2006) it could also be argued that students benefit more from observational tasks opposite to their preferred strategies. Therefore, we explore in the current study how writing preference influences the effect of instructional method on academic writing performance.

Our research design is inspired by Raedts and colleagues and is in line with previous research comparing observational learning with learning by doing. Learning by doing typically includes studying a theoretical part followed by the participants executing several pre-structured writing tasks. In observational learning participants are confronted with the same theoretical part, but instead of executing writing tasks themselves, they observe models performing these tasks, and are asked to reflect upon the behavior of the models (see for example, Braaksma, Rijlaarsdam, & Van den Bergh, 2002; Couzijn & Rijlaarsdam, 2004; Raedts, 2008).

This is also what we have done in the current study. We compared observational learning with learning by doing. All participants studied a theoretical section on effective writing strategies. Afterwards, participants in the observational learning condition observed peer models that were thinking aloud while executing several (pre-) writing tasks. The observers were confronted with both a weak and strong model and they had to reflect on why they thought one model was better than the other. They did not actually write themselves. Participants in the learning by doing condition executed themselves the tasks that the models in the observational learning condition were performing.

The task the participants either observed or executed in the current study was writing an introduction to an empirical research report based on index cards (summaries of scientific articles). This task was new to the participants. Empirical research reports and literature reviews are among the most common types of reports students have to write during their studies (Froese, Gantz, & Henry, 1998). In both text types, writers have to define and clarify the problem, summarize previous investigations and they have to identify relations, contradictions, gaps and inconsistencies in the literature. Compared to the writing tasks in previous research, such as for example Kitsantas and Zimmerman (2002), the academic writing tasks in Raedt's studies and the current study, are more complex and extensive, since writers have to combine multiple source texts, and the texts do not have a pre-arranged structure.

Granello (2001, p. 293) suggests that what seems to be missing in higher education is "a formalized, intentional, and well-grounded mechanism designed to teach students how to critically evaluate and synthesize the material they have collected into cognitively advanced reviews of the literature." This claim is supported by Green and Bowser (2006) who argue that university faculty often assumes that these skills are present in students, while the students might not be able to effectively evaluate and synthesize literature and have not received direct instruction to acquire these skills.

A difference between the literature review in Raedts and the introduction posttest in this study is that participants in our study had to make sure that the synthesis of the literature logically led to the (already provided) research question and hypotheses of the paper they had to write. Furthermore, our posttest, the introduction to an empirical research report, was part of the existing regular course program: the paper the participants had to write was part of the assessment of the course Dutch for Academic Purposes and a methodology course. This way we could test whether the observational learning approach is beneficial in an existing course.

As mentioned earlier, a learner characteristic we want to take into account is writing preference. In the current study four of the six observation exercises focused on more planning-like activities, such as organizing content, planning the structure of the text, and combining information in a paragraph. Students with a planning preference therefore might benefit more from observation as a pre-writing activity than students with a revising preference since most of the observations are based on models performing pre-writing planning activities.

To summarize, in this study we investigate what the effect is of instructional method and writing preference on academic writing performance in learning to write a large and complex writing assignment, namely the introduction of a research paper. We hypothesize that observational learning leads to higher academic writing performance than learning by doing (H1). We expect writers with a planning preference to perform equally well overall as writers with a revising preference. However, we hypothesize that writing preference mitigates the effect of instructional method (H2).

Method

Participants

The participants were recruited from the course Dutch for Academic Purposes, which is an obligatory course for undergraduate students Communication and Information Sciences (n = 211) at Tilburg University. The course offers an introduction to academic writing. The role of writing in academics and different types of texts are discussed. Writing an academic text at university level was new to all participants. Only students who took the course for the first time, were present for both sessions, were Dutch native speakers, and filled out the writing styles questionnaire were included in the analysis. This resulted in a final sample size of 145 participants (male = 54, female = 91). The sample consisted of two types of undergraduate students: 73 first-year bachelor students with a background in preparatory university education (*M_age* = 18.5, *SD* = 1.25) and 72 students enrolled in a pre-master program[1] with a background in higher vocational education (*M_age* = 22.5, *SD* = 1.64).

The participants were divided into nine tutorial groups. They first enrolled in a group by selecting a particular time slot in accordance with their teaching schedule, after which the groups were randomly assigned to one of the conditions: five groups were assigned to the learning by doing condition, four groups to the observational learning condition. Participants were unaware of the conditions when they enrolled in the tutorial groups. This resulted in 81 participants who took part in the learning by doing condition, and 64 participants who took part in the observational learning condition. An overview of the characteristics of the participants per condition can be found in Table 1.

Table 1.
Sex and mean age in years (SDs) of the participants per condition

	Sex		Age		Educational Background	
Instructional Method	Male	Female	M (SD)		Bachelor	Pre-master
Learning by Doing	30	51	51	20.5 (2.46)	43	38
Observational learning	24	40	40	20.5 (2.45)	30	34

Participants in both conditions were comparable in terms of gender (χ^2 (1) = 0.003; *p* = .95), age (*t*(139) = 0.09; *p* = .93) and educational background (χ^2 (1) = 0.55; *p* = .46).

Design

The general design was a 2 (instructional method: observational learning, learning by doing) x 2 (writing preference: plan, revise) design. In the observational learning condition participants were presented with videos in which student-actors were performing writing tasks while thinking aloud. In the learning by doing condition students were confronted with more traditional writing exercises. Prior to the sessions participants filled out a writing style questionnaire. Based on this questionnaire they were either labeled as a planner or reviser. In the posttest the effect of instructional method and writing preference on academic writing performance was measured. An overview of the design can be found in Table 2.

Table 2.
General Research Design

Lecture	Duration (minutes)	Measurement
-	-	Writing style questionnaire Language proficiency test
3	50	Session 1
4	50	Session 2
-	-	Posttest: introduction to an academic paper

Procedures and Material used in the Sessions

Content of the sessions. The sessions took place during the first two tutorials of the course Dutch for Academic Purposes. The participants had no prior knowledge on the subject and no experience with writing an academic text, since these tutorials took place in week 3 and 4 of their study program at the university. In week 2 all participants watched a video lecture on the aims of a scientific article and the different components of an article: abstract, introduction, method, results, discussion, conclusion and references. The tutorials in the third and fourth week of the course program were carried out by four instructors, under supervision of the first author who was one of them. Each instructor was assigned to both an observational learning group and a learning by doing group. A week before the sessions the procedures, a detailed lesson plan and a presentation that contained the exercises were discussed in a meeting with the instructors.

The observational learning condition consisted of two sessions in which six videos were shown (three in each session). Each session lasted 50 minutes. In the videos the participants saw two peer models writing an introduction to an academic paper based on four index cards. All index cards contained a summary of a scientific article that the models had to include in their introduction. Each video focused on a certain aspect of writing an introduction to an academic paper (Table 3). The content of the videos was based on literature on effective and non-effective writing strategies (e.g., Van Weijen, 2009; Graham & Perin, 2007) and suggestions from a study by Raedts, Daems, Van Waes, & Rijlaarsdam (2009). Graham and Perin (2007) found that teaching students strategies for planning, revising and editing (strategy instruction) was a powerful method. They also argued that students should be taught strategies and procedures for summarizing reading material, since this improves their ability to concisely and accurately present this information in writing. Therefore, we included these types of activities in the exercises and videos.

The models in the videos were student actors who had received a script for each exercise and had been instructed to think aloud during the exercise. The script contained concrete instructions for sentences to type and remarks to make while thinking aloud. An example of an instruction for the strong model is: ‘Indicate that the opening paragraph should be about the increased attention there is for students nowadays and their ICT use. Type a bullet point with the keywords for the opening paragraph.’ The student actors were also allowed to give their own input for the exercise, to make sure that the clips were natural and convincing. Two models were used, because observing multiple models increases the likelihood that students will view themselves similar to at least one model (Schunk, 1987). In line with previous research (Groenendijk, Janssen, Van den Bergh, & Rijlaarsdam, 2011; Raedts et al., 2008; Couzijn & Rijlaarsdam, 2004) one of the models used effective strategies to complete the assignments (strong model), the other model used counterproductive strategies (weak model). An effective strategy used by the strong model was comparing and contrasting the studies on the index cards, while the weak model read the studies on the index card without relating the methods and results of those studies. In all the videos the same student actor ("Anne") represented the strong model, and another student actor ("Kristel") the weaker model. The videos were only available for the participants during the sessions.

Table 3.
Content of the Videos in the Observational Learning Condition

Session	Video	Content	Duration
1	1	Reading, selecting, organizing and paraphrasing the information on the index cards	50 minutes
	2	Planning content and main structure	
	3	Organizing the body of the introduction (relating the content information from the different sources)	
2	4	Zooming in on the paragraphs (combining studies within a paragraph)	50 minutes
	5	Adding an opening to the introduction and writing a sentence at the end of the (last) paragraph to bridge the gap to the research questions	
	6	Revising the text at word, sentence and text level.	

The videos were recorded with Camtasia, which allows simultaneous, picture-in-picture recording, and edited with iMovie. Each fragment contained a recording of the model working on the computer, the model's voice and the computer screen the model was working on in Word. By showing a recording of the models, the participants could observe the exact activities of the model, especially when the models were not writing, such as reading and scrolling through the index cards, pausing etc. The recording of the voice gave the participants insight into the thought processes and writing strategies of the models since the models were thinking aloud. The recording of the computer screen gave participants the opportunity to follow the on-screen writing activities of the models, such as typing, deleting and revising text. The length of the videos varied from five to thirteen minutes. Figure 1 illustrates one of the videos. In this fragment weak model Kristel is attempting to write an attractive opening sentence.

The learning by doing condition also consisted of two sessions of 50 minutes each. Participants were in both sessions instructed to write an introduction to an academic paper, based on four index cards. The first session contained five short and pre-structured exercises that led up to a complete introduction. The participants had to complete the exercises under time pressure. The exercises were closely based on the contents of the videos in the observational learning condition as outlined in Table 4 (in the next section the similarities between the two conditions will be discussed in more detail). The second session consisted of one longer, less structured exercise in which the participants were instructed to write an introduction to the topic provided using the information on four index cards. Details of the exercises in the learning by doing sessions can be found in Table 4.

Figure 1.
Screenshot of a fragment with weak model Kristel

Translation of text in Figure 1: “Learning and technology. It is impossible to imagine our society without ICT.”

Similarities and differences between conditions. The conditions were as similar as possible. In the first week of the course all participants watched a video lecture that contained information on the different components of an introduction (opening, literature review, bridge to research question and hypotheses). In both conditions we told the participants that the goal of the exercises during the sessions was to prepare them for writing an introduction to a certain topic (the use of ICT in higher education, the role of gestures in non-verbal communication) by synthesizing information of several academic sources. They were also told that they could use this method of working when writing the introduction of the first paper for the course.

The content of the exercises in both conditions was also similar. Videos 1, 2, 3 and 4 (Table 3) of the observational learning condition correspond with exercises 1, 2, 3 (Table 4) of the learning by doing condition, video 5 is matched with exercise 4, and video 6 with exercise 5. The activities the models performed in the videos of the observational learning condition were the same as the exercises the participants of the learning by doing condition had to execute during the first session.

Table 4.
Content of the Exercises in the Learning by doing Condition

Session	Exercise	Instruction	Time (minutes)
1	1	Read index card 1 and 2 and write a paragraph in which you describe the most important/relevant results. Refer to the sources by using APA standards.	10
	2	Read index card 3. Add the information of the index card to the paragraph you have written in exercise 1.	5
	3	Read index card 4. Add the information to the paragraph of exercise 1 and 2. Divide your text into two paragraphs, if necessary.	5
	4	Read the instructions once more. Add an opening to the introduction and write a sentence at the end of the (last) paragraph in which you bridge the gap to the research questions.	10
	5	Revise your text on text level (structure, composition) and sentence level (spelling, vocabulary, grammar).	5
2	6	Write an introduction to the topic provided. Use the information on the index cards.	50

Participants in both conditions were provided with index cards. As mentioned earlier, each index card contained a summary of an academic paper related to the topic of the introduction. The use of index cards speeded up the reading process and we could make sure all the students received exactly the same relevant information on the topic. On the index cards the following information could be found: the full reference of the article; the research question, the type of research and data; a summary of the most important findings of the study; and a quote from the original article. The first and third index card represented a similar viewpoint. The second and fourth index card contained a viewpoint that was opposite from index card 1 and 3. In Figure 2 an example of one the index cards can be found. The index cards the participants in the observational learning condition received in the two sessions were the same as the index cards the participants of the learning by doing condition used in session 1. Participants in the learning by doing condition were presented with four additional index cards for the second session.

Apart from the index cards we also provided the participants in both conditions with a table that contained effective writing strategies (Table 5). These strategies matched the strategies the 'stronger' model used in the videos. Participants in the observational learning condition had to indicate in the table which strategies they saw the models use in the videos.

Figure 2.
Example of an index card.

Reference

Bennett, S., Maton, K., & Kervin, L. (2008). The 'digital natives' debate: A critical review of the evidence. *British journal of educational technology*, *39*(5), 775-786.

Details about the research

Research theme: analysis of the digital natives debate, with the following research questions:

1. Do students nowadays possess extensive ICT knowledge and skills?
2. Do students have specific learning preferences that are different from earlier generations, because of their experience with ICT?

Type of research: literature review

Data: scientific articles

Results

1. ICT knowledge and skills
   - Part of today's youth has extensive ICT knowledge and skills and uses these skills for information gathering and communication..
   - However, there is also a large group with less access to technology or with less skills than is sometimes assumed.
   - It is dangerous to generalize a whole generation: no room for individual differences between young people or between different age categories.
2. Learning preferences
   - Research into learning strategies shows many individual differences in learning preferences. Students adapt their strategy, dependent on the task.

Conclusion

Not much empirical evidence to support the concept of *digital natives*. ICT can be important, but the situation is much more complex and unclear than expected. More empirical research is necessary to gain insight into the characteristics of students nowadays and the implications for education.

In the learning by doing condition participants used the table to prepare themselves for the exercises. After completion of the exercises, they also had to indicate in the table which strategies they used during the exercises. We included the table in both conditions so the participants were equally exposed to effective writing strategies and in order to include a comparable reflective activity in both conditions.

Procedure in the learning by doing condition. The first session took place during the first tutorial in a computer room. Each participant had access to a computer. The participants received a handout that consisted of an introduction to the assignments, a table with effective writing strategies and corresponding actions, four index cards and an appendix containing basic APA reference guidelines.

Table 5.
Writing Strategies presented to the participants in both conditions

Effective strategies	Corresponding actions
Reading important information	Read the research question and the results.
Organizing the information	Identify general differences and similarities.
Paraphrasing	Put important information into your own words.
Planning	Plan the content and structure by organizing key concepts.
Connecting the content	Identify the differences and similarities between studies in detail.
Combining studies	Connect the content of the different studies in your own text. Use connective words/sentences to mark the relations explicitly.
Adding a quote (optional)	Add a quote using APA guidelines to support an important point.
Revising the text	Check your own text on text level (structure, logic) and sentence level (grammar, vocabulary, spelling).

The instructor explained to the participants that they would execute five small writing exercises in which they would work on an introduction to an academic paper based on four index cards. The participants were given ten minutes to read the instructions, the effective writing strategies and the index cards. Subsequently a presentation containing the first exercise was displayed on a screen. At the end of the exercise an alarm bell went off and the instructor presented the next exercise. At the end of all five exercises the participants were asked to upload their introduction onto the electronic learning environment and to indicate in a table (see Table 5) on the last page of the hand-out which writing strategies they applied during the exercises. The table was the same as the one they had read before doing the exercises. At the end of the tutorial the instructor collected all the handouts. At the start of the second session an example of a good introduction was discussed. This example was the introduction that the strong model ("Anne") wrote in the observational learning videos.

The second session took place during the second tutorial in the same computer room. Participants received another handout that consisted of the same elements as the first one. Participants were instructed to write an introduction to the topic on the handout by using the information on the index cards. They were told that the introduction should contain an effective opening, a synthesis of the academic literature and a bridge sentence to the (already given) research question and hypothesis. The text should also be perfectly written, which meant no spelling or grammar mistakes. They had thirty minutes to complete the assignment. Afterwards, the participants indicated in a table which writing strategies they had applied during the exercise and they had to upload their text onto the electronic learning environment. The instructor collected all the handouts at the end of the tutorial.

Procedure in the observational learning condition. The first session took place during the first tutorial and the second session during the second tutorial, in a regular classroom with a computer for the instructor and a beamer. The procedure was identical for both sessions.

The participants received a handout that consisted of an introduction to the observation exercises, the same four index cards as in the first session of the learning by doing condition and three observation exercises: exercise 1, 2 and 3 for the first session and exercise 4, 5 and 6 for the second session (see Table 3). We told the participants that the goal of the exercise was to learn how to write an introduction to a certain topic by using the information of several academic sources and that they would learn that by observing participants who were completing writing tasks. The participants were unaware of the fact that the models were student actors. We told the participants they could use the strategies they observed while writing the introduction of their first paper for the course. The participants were instructed to carefully read the introduction to the exercises and the information on the index cards. Then the instructor started the first video. There was room on the handout for taking notes while watching. Once the video ended, the participants were given five minutes to answer the following questions: Which differences between the two writers did you observe? Who do you think is the better writer and why? What did the other writer do that made you think she was the lesser writer? The instructor would then start the next video. At the end of the last video of the session and accompanying questions, participants indicated in a table which strategies they saw the models use. This table was identical to the one used in the learning by doing condition (see Table 5).

Measures

Language Proficiency. Prior to the sessions, participants took a test on grammar, spelling and punctuation, and structure in order to check for possible initial differences in participants' language proficiency. The test was developed by the Language Centre of Tilburg University and has been used for over a decade as a diagnostic instrument for undergraduate students at Tilburg University. Grammar was tested with 25 items, containing congruency problems (8), verb conjugations (5) and endophoric expressions (12). Spelling and punctuation were tested with forty items on the spelling of verbs (20) and nouns (13), and the use of punctuation in sentences (7). Structure was tested with ten items on organizing sentences (4), the use of conjunctions (3), and structuring paragraphs (3). Per item one point could be scored, resulting in a possible minimal score of zero, and a maximal score of 25 for grammar, 40 for spelling and punctuation, and 10 for structure.

Writing Preference. Prior to the sessions participants were asked to fill out a questionnaire on writing styles created by Kieft, Rijlaarsdam, & Van den Bergh (2006) in order to determine their writing preference. We chose this particular questionnaire because it has been tested and used in writing research extensively (e.g., De Smet, Brand-Gruwel, Leijten & Kirschner, 2014; Kieft, Rijlaarsdam & Van den Bergh, 2008; Tillema, 2012). The writing style questionnaire measures reported degrees of planning and revising styles and consisted of 36 items: thirteen items reported planning-type behavior, twelve items reported revising-type behavior and the remaining eleven items were fillers. According to the writing style questionnaire planning-type behavior entails pre-writing activities, such as making a text schema and writing a polished first draft. The definition of revising-type behavior is twofold: it focuses on the tendency to rely on revision, and on how revisers use text production as a means to arrive at a content plan (Tillema, 2012).

In the questionnaire participants had to indicate on a five-point-scale how much they agreed with each item (1 = not at all, 5 = very much). An example of a planning item was 'Before I start writing, I want to be clear on which information to put in the text. Therefore, planning is important to me.' An example of a revising item was 'When I finish a text, I usually need to read through it carefully, to check if there is no superfluous information in it.' All the items, organized by dimension, can be found in the Appendix (taken from Tillema, 2012). The items in the actual questionnaire were presented in Dutch and in random order. The items on planning were summarized into one planning score (Cronbach's alpha = .65) and the items on revising into one revising score (Cronbach's alpha = .60). Even though these reliabilities are relatively low, they are comparable to those in previous research (e.g., Tillema, 2012, respectively .72 and .64, and De Smet, Brand-Gruwel, Leijten, & Kirschner, 2014, respectively .71 and .63). Based on their responses, participants received a mean score for both planning and revising.

Academic Writing Performance. To measure the participants' academic writing performance after the sessions the first author scored the introduction section of the first paper the participants had to write for the course Dutch for Academic Purposes. In this paper participants described an experiment on the effects of the use of adjectives on the perceived attractiveness of an advertisement. The participants received four index cards that looked similar to the index cards used in the sessions. They contained the full reference of the article; the research question, the type of research and data; a summary of the most important findings of the study; and a quote from the original article. The studies on index card 1 and 3 showed similarities in their results, while the study on index card 2 displayed an opposing viewpoint. The study on index card 4 added an extra factor that could be of influence on the other three studies. The participants were instructed to write an attractive and suitable opening for their introduction. It was explained that the opening should draw attention of the reader, should not be too formal, nor too informal, and should introduce the topic of the text. These instructions were based on the video lecture the participants watched in week 2. They were also instructed to include all four index cards in the body of the introduction and to make sure that the introduction would lead to the research question and hypotheses in a logical manner. This instruction was identical for all groups.

To measure academic writing performance, the texts were analyzed on the structure of the argument. This dimension was scored on six items: (1) quality of the opening sentence/paragraph, (2) similarity between findings in the studies on index card 1 and index card 3,

(3) contradiction between findings in the studies on index card 1/3 and index card 2, (4) connection between the study on index card 4 and the studies on index cards 1, 2 and 3, (5) paragraph structure, and (6) structure in general (e.g., connective words). For each item zero, one or two points could be appointed which resulted in a possible maximal score of twelve points.

Personal information in the documents was removed by a teaching assistant. Subsequently the texts were scored by the first author, who was unaware of the identity of the respondent and blind to the experimental condition. All the texts have been rescored by two trained student-assistants who each scored half of the papers (interrater-reliability respectively Pearson's r = .76 and .75). The texts were scored using a codebook that included three examples of each of the scores per category possible.

Statistical Analyses

The score on the posttest has been evaluated with an ANCOVA with Instructional Method (*learning by doing, observational learning*) and Writing Preference (*plan, revise*) as the independent factors, and with Educational Background (*bachelor, pre-master*) as covariate, to control for possible effects of previous education.

Results

Language Proficiency and Writing Preference

Before presenting the results of the effect of instructional method on academic writing performance, the general findings concerning initial language proficiency and writing preference are reported.

Initial Language Proficiency. Initial language proficiency was assessed with a grammar, spelling and punctuation, and structure test. Table 6 displays the mean scores on these tests. There were no significant differences between the conditions in scores on grammar ($t(140)= 0.94$; $p = .35$), spelling and punctuation ($t(140)= 0.07$; $p = .94$), and structure ($t(140)= -0.77$; $p = .45$). Both groups can, thus, be assumed to be equivalent concerning language proficiency.

Table 6.
Mean Scores on grammar, spelling and punctuation and structure (SDs) per condition

	Grammar *M (SD)*	Spelling *M (SD)*	Structure *M (SD)*
Instructional Method			
Learning by Doing	18.25 (3.07)	32.73 (3.20)	5.57 (1.29)
Observational learning	17.72 (3.55)	32.69 (3.36)	5.74 (1.33)

Writing Preference. Based on the writing style questionnaire participants received a mean score for planning and revising. Participants who scored higher on planning than on revising were labeled Planners and participants who scored higher on revising than on planning were labeled Revisers. This resulted in 38 Planners and 120 Revisers (see Table 7 for the distribution over conditions). The proportion of participants with a planning (24.1%) and a revising preference (75.9%) appears to be in line with Torrance, Thomas and Robinson (2000). In their longitudinal study, they found that in 23.5% of the 715 essays they analyzed students used a detailed planning strategy, which included activities such as using an outline and one or more idea exploration activities. In the remaining essays students used either mixed or mainly revising strategies, in which ideas were allowed to develop during writing. However, it is worth noting that both in our study and in Torrance et al. (2000) writers with a planning preference also used revising strategies and vice versa: the writing preferences are not mutually exclusive, although in general one of the preferences is more dominant.

Table 7.
Number of Planners and Revisers per condition

	Planners	Revisers
Instructional Method		60
Learning by Doing	21	
Observational learning	17	47

There were no significant differences between the instructional method conditions in writing preferences, ($\chi2$ (1) = 0.01; p = .93). Both conditions can, thus, be assumed to be equivalent concerning writing preferences.

Academic Writing Quality

We investigated the effect of instructional method and writing preference on academic writing quality controlling for educational background. The covariate, educational background, was significantly related to academic text quality, $F(1, 140) = 4.97$, $p = .027$, $\eta_p^2 = .034$. After controlling for educational background, the ANCOVA revealed no main effects for instructional method, $F(1, 140) = 0.40$, $p = .53$, or writing preference, $F(1, 140) = 1.40$, $p = .24$. An overview of the mean scores for academic writing performance for participants with a planning and revising preference in both conditions is presented in Table 8.

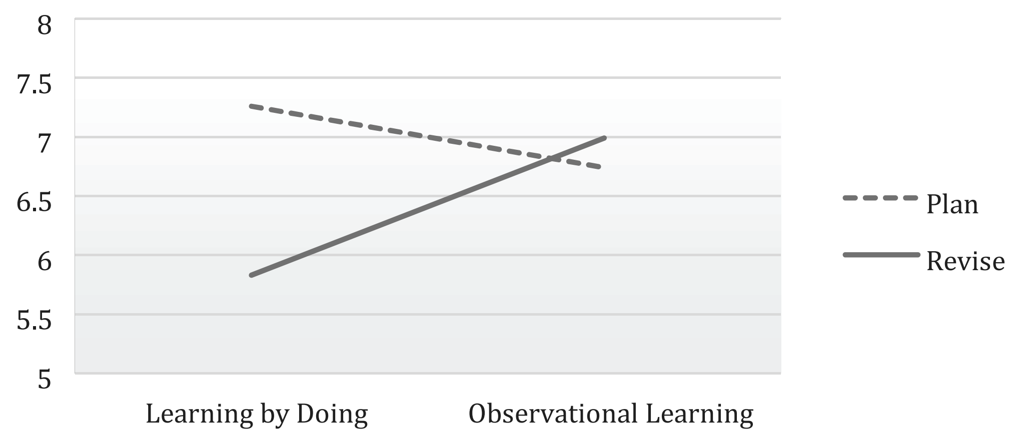
Table 8.
Academic Writing Performance in relation to Instructional Method and Writing Preference (minimal score 0, maximal score 12)

	Learning by doing Mean (*SD*)	Observational learning Mean (*SD*)	Total Mean (*SD*)
Revising preference	5.83 *(2.78)*	6.96 *(2.38)*	6.33 *(2.66)*
Planning preference	7.14 *(2.33)*	6.94 *(3.21)*	7.05 *(2.72)*
Total	6.17 *(2.72)*	6.95 *(2.60)*	

There was an indication for an interaction between instructional method and writing preference, $F(1, 140) = 2.78$, $p = .097$, $\eta_p^2 = .020$. Simple effect analyses revealed a significant effect of instructional method for participants with a revising preference, $F(1, 140) = 5.16$, $p = .025$, $\eta_p^2 = .036$. Revisers in the observational learning condition scored significantly higher ($M = 6.96$) than revisers in the learning by doing condition ($M = 5.83$). For participants with a planning preference no effects of instructional method were found, $F(1, 140) = .36$, $p = .55$.

Simple main effects tests were also conducted holding writing preference constant. Results from these analyses showed that in learning by doing participants with a planning preference scored significantly higher than participants with a revising preference, $F(1, 140) = 4.62$, $p = .033$, $\eta_p^2 = .032$. In the observational learning condition we found no effect, $F(1, 140) = .112$, $p = .74$. Figure 3 provides a graphical representation of the mean scores on academic writing performance.

Figure 3
Academic writing performance in relation to instructional method and writing preference



Discussion

In this study we investigated the effect of instructional method (observational learning versus learning by doing) and writing preference (planning versus revising) on academic writing performance. Our aim was to explore whether observational learning is an effective tool for a large and complex writing assignment, taking writing preference into account. It is worth emphasizing that this study was executed in an ecologically valid situation: in an existing undergraduate course in a regular study program with an actual assignment as posttest.

In the current study no effect of instructional method was found. Contrary to what we expected, participants who learned by observing models did not outperform participants who learned by doing. The two methods seemed equally effective. This does not correspond with the findings by Raedts et al. (2007) who found clear effects of instructional method. Note that our sample was comparable to Raedts et al. and our statistical analysis was sensitive enough to find statistical differences under certain circumstances, indicating that the participants did learn from the sessions. Furthermore, the design of the current study closely followed that of Raedts et al. A minor difference was that we included a form of self-evaluation and reflection in the learning by doing condition. Participants who learned by doing indicated in a table containing effective strategies which ones they had used while performing the tasks. Perhaps this has reduced the difference between observational learning and learning by doing. However, this was only a very small part of the sessions, so we are not sure whether this can account for the differences between the findings.

As expected, we found no effect of writing preference. Planners and revisers performed equally well. This is in line with Galbraith and Torrance (2004) who stated that there is no clear evidence that a certain writing preference leads to a higher writing performance.

We did find some evidence for our second hypothesis: writing preference mitigates the effect of instructional method. Even though we found no overall effect of instructional method on performance, our data revealed different patterns for participants with a planning and a revising preference. Observational learning appeared to be somewhat more beneficial for students with a revising preference. Revisers in the observational learning condition seemed to link the content of the index cards better and they were more successful in adding a suitable opening paragraph and creating a logical bridge to the research questions than revisers in the learning by doing condition. For planners no effect of instructional method was found. This does not appear to be in line with Braaksma, Rijlaarsdam and Van den Bergh (2008) who found that students learned more from a writing course that was adapted to their writing preference. Since most of the observational activities focused on pre-writing planning activities one would expect planners to benefit more from observational learning. In accordance with Galbraith et al. (2006), it could be argued, however, that students with a revising strategy were confronted with pre-writing planning strategies new to them, which could lead to them experimenting with new, more effective strategies while writing the posttest introduction.

Our results also seem to suggest that planners benefit more from learning by doing than revisers. In the learning by doing exercises there is less explicit focus on acquiring pre-writing strategies. Planners tend to use (some of) these strategies anyway. For example, Tillema (2012) found that students who reported a higher degree of planner type behavior on the writing style questionnaire applied more planning activities at the start of task execution. Van Weijen (2008) found a correlation between planning and text quality: participants who have a high probability of occurrence for planning at the start of the writing process, on average, are likely to have written good quality texts. This might explain why planners outperform revisers in learning by doing. This seems to suggest that the more traditional writing instruction suits planners more than revisers. Since the latter group, in our study, represents the majority of the students, it could be argued that observational learning is a more suitable instructional method for learning how to write an academic text since planners and revisers performed equally well when learning by observation. However, it could also be argued that planners can benefit more from observational learning than in the current study when the models focus on different strategies, such as more extensive revision strategies. Combining observational learning and learning by doing could also be an effective option: applying the strategies acquired through observational learning in repetitive, deliberate practice in actual writing could help perpetuate the effects.

More research is needed to fully understand the interplay between writing preference, instructional method and writing performance. For instance, in the current study we have no information on which strategies the students actually used while writing the posttest: the writing style questionnaire is a self-reported measure. Depending on which strategies participants actually apply, the effect of instructional method could be mitigated. Also, it should be noted that perhaps our method of dividing the participants into revisers and planners has been of influence. The participants received a mean score for both planning and revising. Participants who scored higher on planning than on revising were labeled planners and vice versa. This resulted in a revising group that was considerably larger (76%) than the planning group (24%). It also means that even revisers in general may apply some planning strategies and planners will also use revising strategies.

In future studies it would be interesting to investigate the actual planning and revising behavior of participants prior to sessions and during the posttest, for example by including keystroke logging in the design. This way more insight can be gained in how observational learning and learning by doing influence the use of writing strategies and how this is related to writing performance. Also, a pretest measure of academic writing competence should be included. This was missing in the current study. However, since there were no a priori differences in language proficiency between experimental groups, we have no reason to assume initial differences have mitigated the effect of the sessions.

It remains an intriguing question what causes observational learning to be at least as effective as learning by doing, particularly for writers with a revising preference. As mentioned earlier, Braaksma, Rijlaarsdam, Van den Bergh, & Van Hout-Wolters (2004) suggest that this is a result from the observers' strong engagement in metacognitive activities. By evaluating the performance of the models and reflecting explicitly on the observed performances, the intention is that observers develop criteria for effective writing. This evaluating and reflecting in observational learning is in most earlier studies explicitly encouraged by asking participants questions about the model's performance, while in learning by doing explicit evaluation and reflection are typically lacking. In the current study we included some evaluation and reflection in the learning by doing condition by asking participants to indicate which strategies they had used during the interventions. However, this was a smaller part of the interventions than the evaluation and reflection part in the observational learning condition. In future studies it would be interesting to further explore the role of self-evaluation and reflection in both observational learning and learning by doing.

Conclusion

We investigated the effect of instructional method (observational learning versus learning by doing) and writing preference on academic writing performance. More specifically, we wanted to explore whether observational learning is an effective tool for a large and complex writing assignment, and we wanted to explore the possible effects of writing preference.

In this study we found no main effects of instructional method and writing preference. This implies that the students learned how to write the introduction of an academic report equally well with both methods, and that we found no general evidence for one writing preference to result in higher writing performance than the other. However, observational learning seemed to be more beneficial for students with a revising preference. Revisers who learned by observation wrote better organized introductions than revisers who learned by doing. Planners performed equally well in observational learning and learning by doing. However, planners who learned by doing seemed to outperform revisers who learned by doing.

Our study suggests that observational learning can be an effective instructional method in learning to write an academic text in which multiple sources have to be combined and there is no pre-arranged structure. It presents interesting opportunities for (online) courses in academic writing that have little opportunity for individual feedback. However, more information on the exact relation between instructional method, writing preference and academic writing performance is desirable.

Note

1. A Pre-Master is a set of bridging courses that prepares students for a Master's program. For more information on the Dutch educational system, see: https://www.epnuffic.nl/en/ publications/find-a-publication/education-system-the-Netherlands.pdf

**

From this chapter, we conclude that there are some positive albeit weak indications that observational learning might present interesting opportunities for learning to write, but clearly more work is needed. In the next chapter, we look at the comparison between methods in more detail, teasing apart the effect that reflection might have. In observational learning of a cognitive skill, learners are often asked to reflect on the models' behavior explicitly and actively, while it is mostly absent in the learning by doing activities. Since reflective activities in general arguably enhance learning, it is possible that reflection mitigates the effects of observational learning. In Chapter 3 we therefore report on a study in which we systematically compare observational learning and learning by doing, both with and without reflection.

Appendix A

*Items in the Writing Style Questionnaire (Kieft et al., 2006; 2008), sorted according to which dimension they measure. *: item is negatively formulated*

Planning
Before I start writing, I want to have it clear which information to put in the text. Therefore, planning is important to me.
If I have to write a text, I spend a lot of time on thinking about my approach.
I always make a text schema before I start writing.
If I have to write something, I jot down some notes, which I work out later.
Before I start writing a text, I write something on a scribbling pad, to find out my opinion about the topic.
* Planning is of no use to me.
* When I start writing, I don't yet have a clear idea of what will be in the text.
Before I start writing, I have a clear picture of what I want to achieve with the readers.
I need to have my thoughts clear before I am able to start writing.
Before I write a sentence down, I already have it in my head.
* When I am writing, I sometimes write down pieces of text of which I know that they are not completely right yet. Still, I prefer to go on writing at that point.
* When I read over my texts, I usually find a lot to improve.
* When I read over my texts, they are sometimes very chaotic.
Revising
* I always start writing straight away: I don't need to know exactly what I will write or how the text will be built-up. That will become clear as I write.
When my text is ready, I read it through thoroughly and make improvements: a lot can still be changed at that point.
During writing I regularly check if my text does not contain any sentences which are incorrect or too long.
While writing my text, I continually ask myself if readers will be able to follow it.
For me, writing is a way to get my thoughts clear.
* I usually hand in my text without checking if its organization is in order.
If I read over my texts, and rewrite my texts, it occurs regularly that I drastically change their organization.
Before I hand in a text, I always check if its build-up is logical.
* I never pay much attention to whether I have forgotten to put any sentences or ideas in a text.
When I rewrite a text, the content usually changes drastically, too.
When I finish a text, I usually need to read through it carefully, to check if there is no superfluous information in it.
I never pay much attention to whether I am satisfied with my texts.
Fillers
I write and rewrite my text sentence per sentence. Only if I am completely satisfied with a sentence, do I proceed with writing.
When I am writing, I find it hard to organize my thoughts.
Only if my text is complete, do I read what I have written.
If finally I have an approximate idea of what to say in my text, the words will flow out of my pen.
When I write, I stop writing after every few sentences to read what I have just written.
I try to write a correct version of my text in one go, so that I hardly have to make any alterations when it's finished.
When I write a text, I find it hard to come up with ideas.
When I am writing, I often find that all kinds of new ideas pop into my head.
For writing tasks, I do not find it very hard to think of arguments to support my point of view.
The texts which I write are usually not very original.
I make sure that every sentence is perfect, before I start with the next sentence.
When my text is finished, the only thing I do is check for language or spelling mistakes.

CHAPTER 3

Reflection in Learning to Write an Academic Text: How Does Reflection Affect Observational Learning and Learning by Doing in a Research Synthesis Task?

This chapter is based on:
Van der Loo, J., Krahmer, E., & Van Amelsvoort, M. (2019). Reflection in Learning to Write an Academic Text. How does reflection affect observational learning and learning by doing in a research synthesis task? *Frontiers in Education*, 4. https://doi.org/10.3389/feduc.2019.00019

Abstract

In this study, we focused on the effect of reflection on different instructional methods, comparing observational learning and learning by doing, in the context of an academic writing task. Our goal was to investigate how reflection and instructional method affect academic writing performance, self-efficacy beliefs and students' satisfaction with the learning activities. In a quasi-experiment, 111 undergraduate students were assigned to either an observational learning or learning by doing condition, with or without reflection. In the observational learning condition students learned by observing a weak and strong model's writing processes. In the learning by doing condition they learned by performing writing tasks. Half of the students reflected on either the models' or their own performance. In our study, reflection did not affect academic writing performance and self-efficacy beliefs, and neither did instructional method. Both reflection and instructional method did influence students' satisfaction with the learning activities. Students preferred learning by doing over observational learning, and reflecting over not reflecting. From this study, we can conclude that in academic synthesis writing the interplay between reflection, observational learning and learning by doing is not evident yet: students seem to perform equally well in all conditions, even though they prefer learning by doing over observational learning, and reflecting over not reflecting.

Introduction

The effectiveness of observational learning as an instructional method in the writing domain has been studied extensively (e.g., Zimmerman & Kitsantas, 2002; Braaksma, Rijlaarsdam, & Van den Bergh, 2002; Rijlaarsdam et al., 2008, Raedts, Rijlaarsdam, Van Waes, & Daems, 2007; Chapter 2 of this dissertation). In observational learning to write, learners watch and listen to someone else, a model, performing a writing task. By thinking aloud, the model may provide insight into their writing process. This way, learners can construct a cognitive representation of the model's behavior and hopefully connect that to their prior knowledge for future recall (Bandura, 1997; Bandura, 2016).

In writing research, observational learning is often compared to learning by direct experience, mostly referred to as learning by doing. In learning by doing, writers learn by performing a writing task themselves. Only a few studies compared the effect of observational learning to learning by doing in learning to write a complex text at university level (Raedts et al., 2007; Chapter 2 of this dissertation). From these studies, no clear view can be established on the effect of observational learning and learning by doing in writing an academic text.

The goal of this chapter is to get a clearer perspective on the two different methods for writing research, by focusing on one specific aspect, namely reflection. In these aforementioned studies, reflective activities were an intrinsic part of the observational learning process. Observers were explicitly encouraged to carry out different (meta)cognitive, reflective activities, by asking them to compare and evaluate the model's performance. However, reflection was not prompted in the learning by doing conditions, and therefore did not necessarily take place, which may have influenced the findings.

In recent research (e.g., Raedts, Van Steendam, De Grez, Hendrickx, & Masui, 2017; Koster & Bouwer, 2016; MacArthur, Philippakos, & Ianetta, 2015), reflection is often a part of curricula directed at promoting self-regulated learning strategies in order to enhance writing performance. These studies suggest a positive correlation between self-regulated learning curricula, including reflective activities, and writing performance. However, because of the interplay between reflection and other instructional activities in these studies, it is hard to determine the exact role of reflection in learning to write.

In the current study, we investigate the role of reflection in learning to write in a more controlled way. We systematically compare observational learning and learning by doing in learning to write a complex, academic text, with a specific focus on the role of reflection in both methods. In addition to writing performance, we also examine the effects on self-efficacy beliefs and the extent to which learners are satisfied with the instructional methods, because we conjecture that reflection may have an impact on those factors as well. Self-efficacy refers to people's beliefs in their capabilities to produce given attainments (Bandura, 1997) and has consistently been found to be correlated to writing performance (Bruning & Kauffman, 2016). The extent to which learners are satisfied with the instructional method, is also included as a measure. Often there is a mismatch between the way that someone prefers to learn and what actually leads to effective and efficient learning (Kirschner, 2017). An early, illustrative discussion of this can be found in a review by Clark (1982) on studies where two or more instructional methods were allowed to interact with student aptitudes to predict enjoyment and achievement. Clark found that students often report enjoying the method from which they learn the least. In the current study, we therefore explore whether the instructional method, with or without reflection, affects the extent to which learners feel satisfied with the method and whether this is related to writing performance and self-efficacy beliefs.

Observational learning in writing research

Even though the effect of observational learning in learning to write an academic text at university level is still unclear, other studies have reported on the effectiveness of observational learning with other types of writing tasks, with different audiences (e.g., Zimmerman & Kitsantas, 2002; Braaksma, Rijlaarsdam, & Van den Bergh, 2002; Rijlaarsdam et al., 2008; Van Steendam, Rijlaarsdam, Sercu, & Van den Bergh, 2010, Groenendijk, Janssen, Rijlaarsdam, & Van den Bergh, 2013; Fidalgo, Torrance, Rijlaarsdam, Van den Bergh, & Álvarez, 2015). For example, Zimmerman and Kitsantas (2002) demonstrated that observational learning led to better performance on a sentence-combining task. While observational learning did not directly affect self-efficacy beliefs, students who learned by observing models constructed sentences with less errors, stated that they were more satisfied with their performance and showed more intrinsic interest in the task than those who learned by doing. In a study by Braaksma et al. (2002), high school students wrote higher quality argumentative texts when they learned by observing models. Fidalgo et al. (2015) found that upper primary students wrote higher quality texts, and applied more structured and goal-focused planning processes after observing and reflecting on a model. Observational learning has also been reported to influence writing processes in poetry writing with high school students, even though this did not result in better poems (Groenendijk et al., 2013), and to positively affect writing revision skills in foreign language learners (Van Steendam et al., 2013).

Braaksma et al. (2002) argue that the effectiveness of observational learning in these studies can be explained by a reduction in cognitive load in observational learning that allows learners to focus on learning instead of executing tasks. Learners gain insight into the writing processes and the emergence of the resulting text, by observing and reflecting on real models thinking aloud, rather than directly engaging in the cognitively demanding writing task itself (Couzijn, 1999; Rijlaarsdam & Couzijn, 2000). In other words, learners can take a step back from the writing task and focus their cognitive effort on the learning task (Braaksma, Van den Bergh, Rijlaarsdam, Van Hout-Wolters, 2004). This explanation implies that reflection may be an important factor in observational learning. However, this complicates the comparison between the methods, since reflection is typically absent in learning by doing in this field of research.

In this chapter, we want to further investigate the effect of instructional method (observational learning and learning by doing) in learning to write a complex, academic text. The learning task in this study involved writing the introduction to an experimental research paper. This constitutes a type of synthesis text: an introduction typically includes a review of the literature in which writers have to synthesize information from multiple sources. They have to summarize previous studies and identify relations, contradictions, gaps and inconsistencies in the literature, which is a cognitive demanding and complex task (Mateos & Solé, 2009).

In the study by Raedts et al. (2007), 144 undergraduate students learned to write a literature review by either observing a weak and a strong model writing a literature review, or by performing writing exercises. The results indicated that students who learned by observation had more extensive knowledge of effective writing strategies. This was however only the case for strategies concerning information gathering and planning of the text, but not for strategies concerning text production and revision. With regards to text quality, the study showed that students who learned by observation wrote texts of higher quality than those who learned by doing: in the observational learning condition, students linked the source material more often, and wrote better-organized literature reviews compared to the students in the learning by doing condition. Students who learned by observation were also better calibrated for self-efficacy than the students who learned by doing: they were better able to predict their scores, while the students who learned by doing were biased towards overestimating their writing performance.

In Chapter 2, we also compared observational learning with learning by doing in learning to write an academic synthesis text. In a quasi-experiment, 145 undergraduate students were assigned to either an observational learning or learning by doing condition. In observational learning participants learned by observing a weak and strong models' writing processes. In learning by doing they learned by performing writing tasks. In contrast to Raedts et al. (2007), no effects of instructional method were found: students who learned by observing models performed equally well as students who learned by doing. No differences were found in organization and overall quality of the text. There was some indication that writing preference influenced the effect of instructional method. Students with a revising preference appeared to write higher quality introductions when learning by observing models compared to learning by doing. For students with a planning preference no differences were found between methods.

The results by Raedts et al. (2007) and our results from Chapter 2 are thus inconclusive on the effect of instructional method on learning to write an academic synthesis text. However, it should be noted that reflection was present, and actively prompted in the observational learning conditions, but absent and not prompted in learning by doing. In the observational learning conditions, students were asked to evaluate and elaborate on the models' actions. Since several studies report positive effects of reflective activities during learning to write on writing performance, it could be argued that reflection possibly confounded the results by Raedts et al. (2007) and our results from Chapter 2.

Reflection in writing research

Reflection engages individuals in exploring their experiences in order to lead to new understandings and appreciations (Boud, Keogh, & Walker, 2005). It requires learners to explicitly attend to actions and performances and carefully process them, which could contribute to higher transfer performance (Wouters, Paas, & Van Merriënboer, 2009). In the current study, we add reflective activities to learning by doing. To the best of our knowledge, this has not been studied within the observational learning domain. However, as we have seen, findings from studies focusing on self-regulated learning, suggest that including reflective activities in instructional programs, with or without the presence of models, might be beneficial to writing performance and might also affect self-efficacy beliefs (see e.g., Graham, McKeown, Kiuhara, & Harris, 2012).

Within the observational learning domain, reflection is directed at the actions and performances of models. Learners are for example asked to monitor the actions of models, which refers to activities in which deliberate attention is paid to some of one's behavior (Schunk, 1983). Subjects are for example asked how models apply certain theories in the writing task (Braaksma et al., 2001). This requires learners to attend selectively to specific actions and cognitive processes. Learners are often also asked to evaluate and elaborate on models' actions or products, by answering questions as 'Which model is doing better?', and 'Explain why the less well model is doing less well'. To our best knowledge, only the study by Braaksma et al. (2001) has explicitly examined the role of reflection in observational learning within the writing domain.

In order to identify which learning activities were effective within observational learning tasks, Braaksma et al. (2001) examined in a post-hoc study the workbooks, pretests and posttests of 84 ninth-grade students who took part in an observational learning study by Couzijn from 1995. In Couzijn's study, the students participated in four lessons on argumentative reading and writing. They were presented with theory which they had to apply in observation tasks. In these observation tasks, models performed reading and writing tasks while thinking aloud. The models gave insight in orientation, execution and text revision activities, while some models also expressed self-monitoring activities, in which they reflected on their own writing process. The students were explicitly asked to monitor, evaluate and elaborate on the models' performance. They were free to choose whether they would focus on the processes or the products of the models. Braaksma et al. found that reflective activities, in particular evaluating the performance of models and elaborating on the product of the models, contributed to argumentative writing skill. Students who evaluated and elaborated on the model's writing seemed to develop criteria for effective texts and writing processes which transferred to their own writing, yielding higher quality writing performance. They found no effects of process-elaboration on text quality which could possibly be explained by the low number of students that actually elaborated on the models' processes.

When reflective activities are added to learning by doing, these reflective activities are of necessity subtly different from the reflective activities in observational learning: in observational learning, reflection is directed at a model, while in learning by doing, these are directed at the performance of the subjects themselves (self-reflection). Self-reflective activities can be divided into self-observation, self-judgment, and self-reactions (Broadbent & Poon, 2015; Zimmerman, 2008). Self-observation focuses on learners' ability to monitor progress towards their goals. Through self-awareness, learners may develop a better and more appropriate control of certain strategies (Chang, 2007; Zimmerman, 1989). In self-judgment, they evaluate their performance by comparing it to a standard or goal, and by self-reactions learners respond to their performance outcomes, for example by seeking to enhance their personal processes during learning (Zimmerman, 1989).

In recent research, these types of reflective activities are often part of curricula directed at self-regulated learning. Self-regulated learning refers to the modulation of affective, cognitive, and behavioral processes throughout a learning experience in order to reach a desired level of achievement (Sitzmann & Ely, 2011). A number of studies have shown that self-regulated learning curricula, including one or more of these reflective activities, are positively related to learning to write with different types of audiences (e.g., MacArthur, Philippakos, & Ianetta, 2015; Graham, Harris, & Mason, 2005; Zimmerman & Kitsansas, 2002, 1999). For example, MacArthur et al. (2015) found that university students who learned to write an argumentative essay through a self-regulated strategy curriculum, including reflecting on one's progress, wrote essays of overall higher quality than students who followed the regular curriculum in which there was only very limited attention for self-regulation strategies. The self-regulated strategy curriculum also led to a significant increase in self-efficacy scores. In a study with young, struggling writers by Graham et al. (2005) a program with self-regulated learning strategies, including self-monitoring, led to longer, more complete, and qualitatively better papers. Self-efficacy was however not affected by the program. Zimmerman & Kitsantas (1999) found that self-regulated strategies, in particular the reflective activity self-observation (monitoring of one's actions), significantly enhanced writing skill, self-efficacy and self-reaction beliefs with high school students.

These studies on self-regulated learning imply that writing performance could benefit from including self-reflective activities, such as metacognitive monitoring and evaluation. These self-reflective activities might also affect self-efficacy beliefs, although the findings on self-efficacy are not consistent. However, since self-reflection in these studies is part of a larger curriculum, it is hard to establish the exact role it plays in learning to write.

Research Questions

From the findings of the studies discussed, in both the observational learning domain and the self-regulated learning domain, we can conclude that reflective activities, with or without the presence of model, may affect writing performance. Reflective activities might also affect self-efficacy beliefs, even though the effects are not consistent. For example, MacArthur et al. (2015) and Zimmerman and Kitsantas (1999) report an increase in self-efficacy scores, while in Raedts et al. (2007) observational learning leads to a better calibration for self-efficacy, but not necessarily an increase in self-efficacy scores. In Graham et al. (2005) and Zimmerman and Kitsantas (2002), however, self-efficacy was not affected.

To gain a better understanding of the role of reflection in learning to write, we will systematically compare observational learning and learning by doing with and without reflection, applied to academic writing. Our research questions are the following. First, how does instructional method influence academic writing performance, self-efficacy beliefs and satisfaction? Second, how does reflection affect academic writing performance, self-efficacy beliefs and satisfaction? Third, what is the interplay between instructional method and reflection? By focusing on reflection, we aim to get a clearer perspective on the two different methods for writing research.

Methods and Materials

Design

In this quasi-experiment we used a 2x2 between subjects factorial design with reflection (yes, no) and instructional method (observational learning, learning by doing) as factors. This resulted in four conditions: observational learning with reflection, observational learning without reflection, learning by doing with reflection, learning by doing without reflection. We measured the effect on three dependent measures: academic writing performance, self-efficacy and satisfaction with the learning activities. Since we found indications in Chapter 2 that writing preference may influence the effect of instructional method, we included writing preferences (planning, revising) as covariates in the design. In the pretest phase, participants completed a writing style questionnaire, in order to establish writing preference. We also benchmarked participants' initial writing competence and language proficiency. This pretest phase was followed by an intervention phase in which participants learned in two sessions how to write an academic synthesis text by either observing models performing this task (observational learning), or by executing the tasks themselves (learning by doing), both types of learning either with or without reflection. In the posttest phase we measured the effects of both instructional method and reflection on the dependent measures. A schematic overview of the design can be found in Table 1.

Table 1.
Schematic Overview of the Experimental Design

Pretest Phase	Intervention Phase		Posttest Phase
Language Proficiency Test Writing Competence assignment Writing Style Questionnaire	Session 1 (in four conditions)	Session 2 (in four conditions)	Academic writing performance assignment Self-efficacy questionnaire Satisfaction questionnaire

Participants

The participants were recruited from an obligatory course on academic writing for first-year undergraduate students in Communication and Information Sciences at Tilburg University. The course took place in the first weeks of the study program, therefore writing an academic text at university level was new to all participants. The study was first conducted in September 2015 and then run again with a new group of students in September 2016 to increase power. The content, materials, measures and procedures were identical in both years.[1] We included in the analyses only students who took the course for the first time, were Dutch native speakers and completed the posttest Academic Writing Performance. An extra four participants were excluded, because it was unclear in which tutorial group they had been enrolled. The final sample consisted of 111 participants (85 women). The average age was 18.7 years (*SD* = 1.7). The average grade they received on their final examination in Dutch Language at secondary school was 6.8 out of 10 (*SD* = 0.69). The participants enrolled into one of four tutorial groups. Each group was randomly assigned to one of four conditions: observational learning with reflection (n = 20), observational learning without reflection (n = 22), learning by doing with reflection (n = 36) and learning by doing without reflection (n = 33). Table 2 presents an overview of the descriptives of the participants per condition.

1 Post-hoc analyses revealed a significant interaction between reflection and year in which the study was conducted for self-efficacy beliefs, $F(1, 98) = 8.18$, $p = .005$, and satisfaction with the learning activities, $F(1, 63) = 4.36$, $p = .041$. Participants in 2015 who did not reflect had higher self-efficacy beliefs and were more satisfied than the participants in 2016 who did not reflect. No other interactions were found (all *p*'s > .12).

Table 2.
Descriptives Participants per Condition

	Observational learning, Reflection	Observational learning, No Reflection	Learning by doing, Reflection	Learning by doing, No Reflection	Total
Age (in years)	18.6	19.0	18.8	18.4	18.7
Gender	6 males 14 females	3 males 19 females	10 males 26 females	7 males 26 females	26 males 85 females
Grade Final Exam Dutch	6.8	7.0	6.7	6.8	6.8

Materials

Videos in observational learning. Five videos have been used in the experiment. The content of the videos was based on literature on effective and non-effective writing strategies (e.g., Van Weijen, 2009; Graham & Perin, 2007) and suggestions from a study by Raedts et al. (2007) and our study from Chapter 2. In each video two peer models were writing an introduction to an academic paper based on three index cards. The index cards contained a summary of a scientific article, which consisted of the full reference to the article, the research question, the type of research and data and a summary of the most important findings of the study. An example of an index card can be found in Appendix I. The first and third index card represented a similar viewpoint. The second index card contained an opposing viewpoint. Each video focused on a certain task in writing an introduction to an academic research paper. These tasks were identical to the tasks the participants in the learning by doing condition had to perform. The tasks can be found in Table 3.

The models in the videos were two student actors who received a script for each exercise and had been instructed to think aloud during the exercise. Two models were used, because observing multiple models has been argued to increase the likelihood that students will view themselves similar to at least one model (Schunk, 1987). In line with previous research (Groenendijk, et al, 2011; Raedts et al., 2007; our study from Chapter 2) one of the models used effective strategies to complete the assignments (strong model), the other model used counterproductive strategies (weak model). The script contained specific instructions for sentences to type and remarks to make while thinking aloud. The student actors were also allowed to give their own input to the exercise, to make sure that the videos were natural and convincing. The strong model started out by reading the research question, the information about type of research and data, and the findings of each study on the index cards and highlighting relevant information, while the weak model skipped important parts while reading and did not highlight relevant information (task 1). The strong model also connected the different studies on the index cards, establishing similarities and differences, while the weak model did not make any connections between the index cards (task 1 and 2). Then the strong model constructed a text schema while the weak model immediately started writing the opening sentence (task 2). The strong model continued with writing the body of the introduction based on the text schema, paraphrasing the information on the index cards. The weak model wrote the body of the introduction by copying most of the sentences verbatim directly from the index cards (task 3). The strong model then added a suitable opening and described the relevance of the study, while the weak model forgot to add the relevance (task 4). Finally, the strong model revised her text at text, sentence and word level, while the weak model only corrected some minor spelling mistakes (task 5).

The videos were recorded and edited with Camtasia, which allows simultaneous, picture-in-picture recording. The strong and the weak model were depicted alternately in the videos. In order to avoid recency and primacy effects, three videos started with the weak model and two videos with the strong model. Each fragment contained a recording of the model working on the computer, the model's voice and the computer screen the model was working on in Word. Figure 1 displays a screenshot of the weak model in exercise 4.

Exercises in Learning by doing. In the learning by doing conditions participants executed five writing exercises in which they wrote the introduction to an academic paper based on the same three index cards as the models in the observational learning videos used. The exercises were identical to the tasks the strong model executed in the observational learning videos (see Table 3). In the first exercise participants were instructed to read the research question, the information about type of research and data, and the findings of each study on the index cards and to highlight relevant information. In the second exercise participants had to reread the index cards and write down the most important similarities and differences between studies. Then they were instructed to make a text schema based on their notes. In the third exercise participants had to write the body of their introduction in which the most important findings of the studies were described. In the fourth exercise, they were asked to add an opening to the body of the introduction and to add a closing paragraph in which they described the scientific relevance of the study. In the fifth exercise, they were asked to check their text at text, sentence and word level.

Reflection Questions. To prompt reflection, participants in the reflection conditions answered three questions about either the performance of the models they observed (observational learning) or about their own performance (learning by doing). Participants in the observational learning condition were asked to answer the following questions: Which differences did you observe between the two writers? Who do you think is the better writer and why? What did the other writer do that made you think she was the lesser writer? In learning by doing participants answered the following questions: How did you handle the last exercise? What went well during the exercise? What would you do differently next time? These questions allowed participants to identify, evaluate and elaborate on writing strategies in both conditions.

Procedure for the Experimental Session

Prior to the experimental sessions all participants were asked to complete the language proficiency test and the writing competence assignment via the university's learning management system. They also filled out the writing style questionnaire online. The experimental sessions all took place during the first two tutorials of the academic writing course and were led by the same instructor (the first author).

Observational Learning without reflection. The sessions took place in a regular classroom. At the start of the first session the instructor told the participants they were going to watch five videos of two models writing the introduction to an academic paper: two during the first session and three in the second session. Then the instructor distributed a handout that contained an introduction to the tasks, and the index cards. The instructor read the introduction out loud while the participants could read the text on the handout. The instructor also explained what the index cards were for. Then the instructor started the first video. The video was projected with a beamer. All participants watched the same screen so the instructor could check whether all participants were looking at the video. If not, the participants were redirected by the instructor to the screen. The first video was followed immediately by the second one. At the end of the first session the instructor collected all the handouts. At the start of the second session the handouts were redistributed, and the introduction was repeated. The participants then watched the three remaining videos in the same way as during the first session. At the end of the second session the instructor collected all the handouts again. The participants spent 55 minutes in total watching the videos.

Observational Learning with reflection. The procedure in the observational learning with reflection condition was identical to the procedure without reflection, but the reflection questions were added. Before the instructor started the first video, the participants were asked to make notes on their handouts while watching the videos. After each video, they had a maximum of five minutes to answer the reflection questions on the handout.

Learning by doing without reflection. The learning by doing sessions took place in a computer room. Each participant had access to a computer. The instructor told the participants that they would be writing an introduction to an academic paper based on three index cards by executing five pre-structured exercises: two during the first session and three during the second session. Then the instructor distributed a handout that contained an introduction to the tasks, and the index cards. The instructor read the introduction out loud while the participants could read the text on the handout. The instructor also explained what the index cards were for and told the participants that they could use the computer to execute the tasks. Then the first exercise was displayed through a beamer on a screen. The participants had 10 minutes to complete the first task. Then the second exercise was displayed. After 15 minutes the participants were asked to upload their work on the first two exercises and to hand-in the handouts. At the start of the second session the handouts were redistributed, and the introduction was repeated. The participants could also open the file with their work on the first two exercises on their computer, and they had five minutes to review their previous work. Then the third exercise was started, followed by the fourth and fifth exercise. They had respectively 15, 10 and 10 minutes to complete the assignments. In total participants spent 55 minutes working on the exercises. At the end of the second sessions the participants were asked to upload their work on all the exercises and the instructor collected the handouts.

Learning by doing with reflection. In the learning by doing condition with reflection the procedure was identical to learning by doing condition without reflection, except the reflection questions were added. After each exercise participants were asked to reflect for a maximum of five minutes on their activities by answering the reflection questions on the handout.

Measures

Pre-test writing performance. To benchmark the participants' initial writing performance, they were asked to write an argumentative text. We avoided administering a pre-test on writing an academic text because of possible learning effects: we wanted the task in the post-test to be new to all participants. For the pre-test, participants therefore were asked to write an argumentative text on the future use of English as a lingua franca at Dutch universities. They were provided with a list of eight arguments pro and eight arguments contra. To make sure the texts were comparable, we asked all participants to take position against the statement. In the assignment, the participants had to combine and integrate arguments. This way the pre-test resembled the post-test as closely as possible. The pre-test was scored on the number of arguments (min. 0, max. 3 points), the organization of the arguments (min. 0, max. 3 points), the quality of the opening sentences (min. 0, max. 2 points) and the conclusion (min. 0, max. 2 points) and the general structure of the text (min. 0, max. 2 points). This resulted in a maximum possible score of 12 points.

Pre-test language proficiency. All participants completed a language proficiency test on grammar, spelling and punctuation, and structure in order to check for possible initial differences. The test was developed by the Language Center of Tilburg University and has been used for over a decade as a diagnostic instrument for undergraduate students at Tilburg University. Grammar was tested with 25 items, containing congruency problems (8), verb conjugations (5) and endophoric expressions (12). Spelling and punctuation were tested with forty items on the spelling of verbs (20) and nouns (13), and the use of punctuation in sentences (7). Structure was tested with ten items on organizing sentences (4), the use of conjunctions (3), and structuring paragraphs (3). Per item one point could be scored, resulting in a possible minimum score of zero and a maximum score of 75 points for the language proficiency test[2].

Writing style questionnaire. Prior to the sessions participants were asked to fill out a questionnaire on writing styles created by Kieft, Rijlaarsdam, & Van den Bergh (2006) in order to determine their writing preference. We chose this particular questionnaire because it has been tested and used in writing research extensively (e.g., De Smet, Brand-Gruwel, Leijten & Kirschner, 2014; Kieft, Rijlaarsdam & Van den Bergh, 2008; Tillema, 2012; Chapter 2 of this dissertation). The writing style questionnaire measures reported degrees of planning and revising styles and consisted of 36 items: 13 items reported planning-type behavior, 12 items reported revising-type behavior and the remaining 11 items were fillers. In the questionnaire participants had to indicate on a five-point-scale how much they agreed with each item (1 = not at all, 5 = very much). An example of a planning item was 'Before I start writing, I want to be clear on which information to put in the text. Therefore, planning is important to me.' An example of a revising item was 'When I finish a text, I usually need to read through it carefully, to check if there is no superfluous information in it.' All the items, organized by dimension, can be found in Appendix III (taken from Tillema, 2012). The items in the actual questionnaire were presented in Dutch and were in random order. The items on planning were summarized into one planning score (Cronbach's alpha = .65) and the items on revising into one revising score (Cronbach's alpha = .60). Even though these reliabilities are relatively low, they are comparable to those in previous research (e.g., Tillema (2012) respectively .72 and .64, and De Smet et al. (2014) respectively .71 and .63). Based on their responses, participants received a mean score for both planning and revising. Since the correlation between the scores was only moderate (Pearson's r = .40), we included both dimensions separately in the design. No differences were found between conditions on planning score and revising score.

Academic writing performance. To measure academic writing performance, we scored the introduction of the first paper participants had to write for the academic writing course. For this writing task, participants were provided with three index cards similar to the index cards used in the sessions. The studies on index card 1 and 3 showed similarities in their results, while the study on index card 2 displayed an opposing viewpoint. Participants were instructed to write an attractive and suitable opening for their introduction, to include all three index cards in the body of the introduction and to make sure that the introduction would lead to the research question and hypotheses in a logical manner. The participants wrote their assignments at home, and handed in the assignment one week after the second experimental session. They had to indicate how much time they spent on the assignments (M = 98.5 minutes). We found no differences between conditions in the amount of time they spent on the assignment (all p's > .71). The texts were scored on six dimensions: (1) opening paragraph (2) similarity between index card 1 and index card 3 (3) contradiction between index card 1/3 and index card 2 (4) mentioning of scientific relevance (5) mentioning of social relevance (6)

2 The language proficiency test is available, in Dutch, on request (please contact the first author).

structure in general (e.g., connective words). For each item zero to two or zero to three points could be assigned which resulted in a possible maximal score of fifteen points. The scoring scheme was derived from to the scoring used in Chapter 2. The texts were scored using a codebook that included three examples of each of the scores per category possible. All texts have been re-scored by a trained student-assistant (inter-rater reliability Pearson's r = .86).

Self-efficacy. Before writing the introduction that was used to measure academic writing performance, participants had to complete a self-efficacy questionnaire. Self-efficacy was measured with ten items. The items were closely linked to the tasks performed in the videos and exercises, in line with Zimmerman and Kitsantas (2002), and were identical to the items used in Chapter 2. Five items were related to writing conventions, which refers to 'accepted standards for expressing ideas in a given language' (Bruning, Dempsey, Kauffman, McKim, & Zumbrunn, 2013, p. 28). An example of such an item is: "I am able to write grammatically correct sentences". The other five items were related to ideation, which refers to the participants' beliefs about their ability to generate ideas (Bruning et al., 2013). An example item is: "I am able to paraphrase information from the index cards." Participants had to indicate on a scale from 0 (not confident at all) to 100 (very confident) with ten-point intervals how confident they felt that they were able to write the introduction. The average score (max. 100 points) of the ten items (Cronbach's alpha = .89) was used as the participants' self-efficacy score.

Satisfaction. After handing in the self-efficacy questionnaire and the academic writing performance assignment, participants received a questionnaire in which we measured the participants' satisfaction with the instructional method, with or without reflection. Participants had to indicate on a scale from 1 to 5 (1 = not at all, 5 = very much) to what extent they agreed with five items. Examples of items are: "The videos / exercises were useful" and "The videos / exercises helped me in writing the introduction". The average score of the five items (Cronbach's alpha .83) was used as the participants' satisfaction score.

Statistical analysis

The scores on the posttest have been evaluated with three separate ANCOVA's with Instructional Method (learning by doing, observational learning) and Reflection (yes, no) as the independent factors and Academic Writing Performance, Self-Efficacy and Satisfaction as the dependent factors[3]. We included the mean scores on Planning and Revising Preference as covariates to control for possible effects of writing preference[4].

3 We performed separate ANCOVA's rather than one MANCOVA, because of missing data in the dependent variable Satisfaction. 71 participants of the 111 participants who handed in the academic writing performance assignments, responded to the satisfaction questionnaire. These 71 participants were equally distributed over conditions.

4 By adding two covariates the model becomes more complicated. To check whether this did not negatively affect the results, we performed a post-hoc analysis without Planning and Revising Preference as covariates. This did not

Results

Initial writing performance and language proficiency

To check for initial differences between conditions we performed an ANOVA on initial writing performance and language proficiency. Initial writing performance was benchmarked by scoring an argumentative text participants wrote before the experiment started. We found no differences between the instructional method conditions, $F(1, 90) = 0.38$, $p = 0.54$, and the reflection conditions, $F(1, 90) = .25$, $p = 0.62$. There was no significant interaction either, $F(1, 90)$, = 1.16, $p = 0.29$. An overview of the mean scores can be found in Table 4.

Table 4.
Mean Scores (SDs) on Initial Writing Performance (min. score 0, max. score 12)

	Observational learning	Learning by doing	Total
Reflection	8.35 (2.09)	7.64 (2.13)	7.88 (2.13)
No Reflection	7.68 (2.11)	7.88 (1.59)	7.80 (1.81)
Total	8.00 (2.10)	7.74 (1.91)	

For initial language proficiency, there were no significant differences between the instructional method conditions, $F(1, 89) = 0.12$, $p = 0.73$ and the reflection conditions, $F(1, 89) = .65$, p = 0.42.There was no significant interaction either, F(1, 89) = .30, $p = 0.59$. Table 5 displays the mean scores on this test. Based on the pretests there is no reason to assume there were a priori differences between experimental groups.

Table 5.
Mean Scores (SDs) on Language Proficiency Test (min. score 0, max. score 75) per Condition

	Observational learning	Learning by doing	Total
Reflection	59.44 (4.02)	59.65 (5.14)	59.6 (4.71)
No Reflection	60.85 (5.09)	59.92 (5.06)	60.3 (5.04)
Total	60.2 (4.61)	59.8 (5.06)	

Academic writing performance

A 2 (instructional method) x 2 (reflection) ANCOVA was calculated on participants' academic writing performance, with planning and revising preference as covariates. An overview of the mean scores for academic writing performance in all conditions is presented in Table 6.

We found no main effects for instructional method, $F(1, 105) = 0.000$, $p = .98$. Participants who learned by doing scored equally well as those who learned by observing models. Also no main effect for reflection was found, $F(1, 105) = 1.85$, $p = .18$. Reflecting on the models' activities or participants' own activities did not affect academic writing performance. No interaction was found either, $F(1, 105) = 1.58$, $p = .21$. Further inspection of the table shows that the

affect the results. We also, after suggestions from a reviewer, performed a post-hoc analysis including pre-test scores, initial language proficiency and grade final examination Dutch Language as covariates. This did not affect the results.

mean scores are relatively close to each other, while the standard deviations are relatively large. This implies large variations between participants within conditions. The covariates, Planning Preference, $F(1, 105) = 1.83$, $p = .18$, and Revising Preference, $F(1, 105) = 0.062$, $p = .80$, were not significantly related to academic writing performance.

Table 6.
Average Academic Writing Performance (SDs) in relation to Instructional Method and Reflection (minimal score 0, maximal score 15)

	Observational Learning	Learning by Doing	Total
Reflection	7.88 (3.37)	8.58 (3.21)	8.33 (3.25)
No Reflection	7.66 (3.74)	7.00 (2.81)	7.26 (3.20)
Total	7.76 (3.53)	7.83 (3.11)	

Self-efficacy

The ANCOVA revealed no main effects for instructional method, $F(1, 102) = 0.309$, $p = .58$, or reflection, $F(1, 102) = 1.227$, $p = .27$. Participants who learned by observing models did not have more confidence in the writing task than those who learned by doing. Reflection did not yield higher scores either. No interaction was found either, $F(1, 102) = 0.265$, $p = .61$. As can be seen in Table 7, there is limited range in the mean scores. All scores are relatively high. Again, there appear to be large variations within conditions. No interaction was found either, $F(1, 102) = 0.265$, $p = .61$. The covariates, Planning Preference, $F(1, 102) = 0.409$, $p = .52$, and Revising Preference, $F(1, 102) = 0.008$, $p = .93$, were not significantly related to self-efficacy.

Table 7.
Self-efficacy Scores (SDs) in relation to Instructional Method and Reflection (minimal score 0, maximal score 100)

	Observational Learning	Learning by Doing	Total
Reflection	70.31 (8.62)	67.73 (10.68)	68.66 (9.99)
No Reflection	66.32 (15.08)	66.46 (10.95)	66.40.26 (12.63)
Total	68.27 (12.37)	67.57 (11.34)	

Instruction Evaluation

To evaluate the instructional activities (learning by doing or observational learning, and reflection or no reflection) we asked participants how satisfied they were with the activities. The covariates, Planning Preference, $F(1, 65) = 0.206$, $p = .65$, and Revising Preference, $F(1,65) = 0.034$, $p = .85$, were not significantly related to the satisfaction score.

The ANCOVA showed significant main effects of instructional method, $F(1, 65) = 31.63$, $p < .001$, $\eta_p^2 = .33$ and reflection, $F(1, 65) = 6.58$, $p = .01$, $\eta_p^2 = .092$. Participants who learned by doing were significantly more satisfied with the learning activities than participants who learned by observing, and participants who reflected were more satisfied than participants who did not reflect. A significant interaction was found between instructional method and reflection, $F(1, 65) = 6.52$, $p = .01$ $\eta_p^2 = .091$.

Simple effects analyses revealed that participants who learned by observing and did not reflect were the least satisfied with the instructional activities. They were significantly less satisfied than participants who learned by doing and did not reflect, $F(1, 67) = 36.84$, $p < .001$, $\eta_p^2 = .36$. For the participants who did reflect, the ones who learned by observation were significantly less satisfied than the ones who learned by doing, $F(1, 67) = 4.79$, $p = .032$, $\eta_p^2 = .067$. Simple effects analyses also showed that participants who learned by observing and reflected were significantly more satisfied than those who learned by observing and did not reflect, $F(1, 67) = 14.248$, $p < .001$, $\eta_p^2 = .18$. For participants who learned by doing no differences were found between those who reflected and those who did not reflect, $F(1, 67) = 0.001$, $p = .97$. Table 8 displays the mean satisfaction scores per condition.

Table 8.
Satisfaction scores (SDs) in relation to Instructional Method and Reflection (minimum score 0, maximum score 5)

	Observational Learning	Learning by Doing	Total
Reflection	$3.51^{b,c}$ (0.38)	3.93^{b} (0.48)	3.71 (0.48)
No Reflection	$2.83^{a,c}$ (0.74)	3.92^{a} (0.46)	3.32 (0.83)
Total	3.13 (0.69)	3.92 (0.47)	

Note. ^{a,c} significant at the .01 level, ^{b} significant at the .05-level

Correlational analyses

To examine relations among the dependent measures, we performed Pearson correlation analyses. Table 9 displays the obtained correlation coefficients. Self-efficacy beliefs appeared to be related to academic writing performance. None of the other relations were statistically significant.

We also explored the correlations within the different instructional methods condition (Table 10) and reflection conditions (Table 11). For learning by doing, and for reflection, there are small but significant correlations between writing performance and self-efficacy.

Table 9.
Correlations among Dependent Measures

	1	2	3
Academic Writing Performance	-		
Self-efficacy Beliefs	.161*	-	
Instruction Evaluation	.090	.091	-

* Correlation is significant at the .05 level (1-tailed).

Table 10.
Correlations among Dependent Measures in Observational Learning and Learning by Doing Conditions

		1	2	3
Observational learning	Academic Writing Performance	-		
	Self-efficacy Beliefs	.108	-	
	Instruction Evaluation	.128	.022	-
Learning by Doing	Academic Writing Performance	-		
	Self-efficacy Beliefs	.204*	-	
	Instruction Evaluation	.028	.133	-

* Correlation is significant at the .05 level (1-tailed).

3

Table 11.
Correlations among Dependent Measures in the Reflection and No Reflection Conditions

		1	2	3
Reflection	Academic Writing Performance	-		
	Self-efficacy Beliefs	.241*	-	
	Instruction Evaluation	-.055	.114	-
No Reflection	Academic Writing Performance	-		
	Self-efficacy Beliefs	.072	-	
	Instruction Evaluation	.145	.051	-

* Correlation is significant at the .05 level (1-tailed).

Discussion

Writing an academic synthesis text is a complex and demanding task for students, since it involves different cognitive activities, such as thinking about content, planning the text, and translating ideas into sentences, that have to be executed simultaneously. In this study, we investigated what would be the most effective way for students to learn to write such an academic synthesis text. We focused on the effect of reflection on two different instructional methods, systematically comparing observational learning and learning by doing, since it is not clear from previous studies what the exact role of reflection entails. Our goal was to investigate how reflection and instructional method affect academic writing performance, self-efficacy beliefs and students' satisfaction with the learning activities. The main findings indicate that instructional method and reflection did not seem to affect academic writing performance or self-efficacy beliefs. Students, however, preferred learning by doing over observational learning, and reflecting over not reflecting.

Observational learning versus learning by doing

With regard to instructional method (observational learning and learning by doing) our results are in line with our previous study in Chapter 2. In these two studies, students who learn by doing and students who learn by observing produce academic texts of equal quality. The texts do not differ in structure or in the coherence of the different sources used in the text. This seems to suggest that it does not matter which method is used to teach students how to write an academic text: as long as they are actively engaged in their classes, they learn. We did find large individual differences within conditions, suggesting that individual characteristics might have an influence on possible effects of instructional method.

Students in both instructional methods also have equal confidence in their performance of the task. In line with Zimmerman and Kitsantas (2002) no direct effects of observational learning were found. Closer inspection of the correlations even revealed a higher correlation between self-efficacy and writing performance in learning by doing than in observational learning, even though the correlation was only small. This is in contrast to Raedts et al. (2007) who found a better calibration for self-efficacy in observational learning. In the current study, all self-efficacy scores were relatively high. Since the task was new to all students and they had not received teacher or peer feedback at that point, it is possible that there was a tendency for students to overestimate their ability, irrespective of instructional method.

An interesting finding is that students prefer learning by doing over observational learning. It is possible that they experienced the learning by doing exercises as more "active" and therefore more useful than merely watching someone else perform a task. Learning by doing is probably also more closely linked to what they expect from a writing course which could have affected motivation for the writing task.

Reflection in learning to write

In the current study, contrary to what we expected, reflection did not affect academic writing performance and did not lead to higher self-efficacy beliefs. Students who reflected did not write better texts than those who did not reflect, while in previous studies in which reflection was embedded in self-regulated-learning curricula, reflection did appear to enhance learning. Similar to instructional method, we found large individual differences within conditions.

It is worth noting that reflection in the different learning conditions was of necessity subtly different: in observational learning the students reflected on the models' performances while in learning by doing they reflected on their own performance, and this difference might conceivably influence the effectiveness of reflection. However, in earlier research (e.g., MacArthur et al., 2015; Zimmerman & Kitsantas, 1999) both types of reflection yielded higher learning performances. We did find some indications that reflection affected self-efficacy. There was a higher correlation between self-efficacy and academic writing performance scores for the students who reflected than for those who did not reflect. This suggests that reflection might lead to more accurate self-efficacy beliefs. It should be noted, however, that these correlations were only small. Importantly, we did find that students preferred reflecting over not reflecting, which implies that students do consider reflection to be a useful learning activity.

How can we explain the differences between our findings and those of previous self-regulated learning research? In MacArthur et al. (2015), the reflective activities were part of a larger curriculum directed at self-regulated learning. Within the curriculum students received other kinds of treatments, such as direct instruction, teacher modeling and group practice over a complete semester in nine classes. These students were therefore exposed more often and for a longer period of time to multiple types of self-regulating strategies, allowing them to accommodate new information or strategies within their mental models several times, which could explain the differences with our study.

In our study, by contrast, we systematically compared reflecting to not reflecting in both observational learning and learning by doing. Since we found no effects of reflection, it could be argued that reflection by itself does not suffice for helping students to learn a complex writing task.

Limitations and further research

The study was conducted in an ecologically valid situation which did allow us to examine the effects of the experimental sessions in real life: the sessions were part of an existing course in academic writing and the task we used to measure academic writing performance was part of the regular assessment for the course. However, this makes it difficult to control for other factors. There was substantial variation in scores within conditions, implying that individual differences between the learners might have mitigated the effects of instructional method and reflection, which complicates establishing an overall effect of learning strategy. For example, Kellogg (2008) argues that domain-specific knowledge on the topic allows writers to focus more executive attention to the writing process, which in turn could yield higher quality texts. The topic of the posttest in the current study was presumably new to all students, but it is possible that some students comprehended the content better than others.

It could also be argued that other learner characteristics have influenced the effects of the learning activities. Affective-motivational constructs, such as academic self-concept, academic interest, and academic anxiety could determine academic effort, such as motivation for the task, and time and effort spent on the task (Schunk, Pintrich, & Meece, 2010; Gogol, Brunner, Martin, Preckel, & Götz, 2017). Even though we did not find differences between experimental groups in time spent on writing the task, it is possible that some students spent more time on preparing for the task.

We also cannot determine with certainty which writing strategies the students used since we did not measure their actual writing process. To determine the actual strategies used, it might be relevant, however, in future studies to include keystroke logging in the design, to gain more insight in how instructional method and reflection influence the use of writing strategies and how this relates to writing performance.

With regards to reflection, it should be noted that the reflection questions in our study were solely focused on the writing process, and not on affective constructs or on the product, which does seem to be the case in self-regulated learning curricula. Interestingly, Braaksma et al. (2001) also found no effect of reflection on the writing process, even though they suggest this was due to the low number of students actually reflecting on the process.

Theoretical implications

The findings of this study contribute to existing research on observational learning and reflection. Several issues are worth noting. First, even though students preferred learning by doing over observational learning, and reflecting over not reflecting, this did not positively affect writing performance. This seems to imply that preference for a specific learning activity does not enhance learning. This is in line with the findings of Clark's review (1982). Learner preferences are typically not correlated or even negatively correlated with learning and learning outcomes, which implies that there is a difference between the way in which someone prefers to learn and that which actually leads to effective and efficient learning (Clark, 1982; Kirschner & Van Merriënboer, 2013).

Secondly, this study implies that reflection does not necessarily increase self-efficacy, but that it might affect the accuracy of self-efficacy beliefs. According to Bandura (1997) self-efficacy can be promoted by modeling. By watching someone else succeed in a task, learners might feel more confident that they can succeed as well, which could lead learners to participate more readily in a task, work harder, persist longer when they encounter difficulties, and achieve at higher levels (Schunk & Zimmerman, 2007). In our study however, modeling did not raise self-efficacy beliefs. We did find indications that identifying, evaluating and elaborating on the writing processes of models, and interestingly, also on one's own writing processes, might lead to more accurate self-efficacy beliefs.

Another issue worth mentioning is that in both this study and our previous one from Chapter 2, we found no differences between learning activities and the effect they have on academic writing performance. This seems to imply that being active in class is sufficient, independent of the learning activities. This is supported by Credé, Roch, and Kieszczynka (2010) who conclude from a meta-analysis of the relationship between class attendance in college and college grades that mere class attendance is strongly correlated with academic performance. According to Credé et al. class attendance is a stronger predictor of academic performance than any other known predictors, including study skills.

It could also be argued that the former issue is related to the large variation in scores we found within conditions. This variation implies that there are considerable individual differences between students when it comes to writing. How to adapt instructional methods in such a way as to meet these differences and to improve learning, is challenging. For instance, Kirschner and Van Merriënboer (2013) suggest that cognitive abilities and prior knowledge should be taken into account when instructional methods are applied, since there is evidence that, for instance, learners with high prior knowledge benefit from other instructional methods than learners with lower prior knowledge.

Conclusion

In this study, we focused on the effect of reflection on different instructional methods, comparing observational learning and learning by doing, in the context of an academic writing task. Our goal was to investigate how reflection and instructional method affect academic writing performance, self-efficacy beliefs and students' satisfaction with the learning activities. In our study, reflection did not affect academic writing performance and it did not increase self-efficacy beliefs, and neither did instructional method. There were some indications that reflection leads to more accurate self-efficacy beliefs. Both reflection and instructional method did influence students' satisfaction with the learning activities. Students preferred learning by doing over observational learning, and reflecting over not reflecting. From this study, we can conclude that in academic synthesis writing the interplay between reflection, observational learning and learning by doing is not evident yet: students seem to perform equally well in all conditions, even though they prefer learning by doing over observational learning, and reflecting over not reflecting.

**

The two single subject design studies from Chapter 2 and 3 were conducted in a 'real life' writing course. This supports the ecological validity of the studies, but it complicates controlling for other factors that might be of influence. In Chapter 4 we therefore attempt to study observational learning and reflection in a more controlled environment (a lab study) and with a less complex writing task (sentence combining).

CHAPTER 4

4

Learning How to Combine Sentences: Effects of Modeling Type and Reflection on Writing Revision Proficiency

This chapter is based on:
Van der Loo, J., Krahmer, E., & Van Amelsvoort, M. *Learning How to Combine Sentences. Effects of Modeling Type and Reflection on Writing Revision Proficiency.* Manuscript submitted for publication.

Abstract

Constructing syntactically complex and correct sentences is a difficult but important task in composing texts. Therefore, it has been argued that teaching students to combine sentences is beneficial for teaching them how to write texts. In this study we investigate which instructional method is most effective for teaching university students to combine sentences by conceptually replicating Zimmerman and Kitsantas (2002). More specifically, we compare the effect of observational learning (learning from observing a coping or mastery model) to learning by doing (by performing writing tasks). In addition, we investigate the role of reflection. In our study, 108 undergraduate students were taught sentence combining by either watching a coping model or a mastery model performing sentence combining tasks, or by practicing with sentence combining tasks themselves. Additionally, half of the students in either instructional method was explicitly asked to reflect on the learning process, while the other half was not. From our findings, we conclude that neither modeling type nor reflection affects writing revision proficiency. Students who observed a model did show an initial increase in their self-efficacy beliefs which disappeared after a practice round. Learning from both the coping and mastery model did lead to a more accurate implementation of the proposed solution strategy and the students preferred observing a model over mere practicing. This did however not lead to higher performance scores. The implications of these findings are discussed.

Introduction

Constructing sentences is an essential part of the writing process. It involves transforming ideas and intentions into written text that makes sense while conforming to the permissible syntax of the language (Hayes & Flower, 1986; Saddler & Graham, 2005). The ability to construct syntactically correct and complex sentences is arguably a core writing skill which facilitates expressing more complex relationships between ideas, such as cause-and-effect-relationships (Coirier, 1996) and which allows writers to express their ideas in a concise manner (Beers & Nagy, 2009). If writers have limited knowledge on how to construct syntactically correct and meaningful sentences, this might negatively affect the writers' abilities to communicate their thoughts and ideas. It might also negatively impact the readability and comprehension of their texts (Saddler & Graham, 2005). It therefore seems beneficial to explicitly teach students sentence construction skills, even at university level (Saddler & Graham, 2005).

A way of teaching students how to construct sentences is using a sentence-combining task, which can be defined as "teaching techniques for splicing together simple sentences to make compound or complex ones" (Andrews et al., 2005, p. 42). In other words, it teaches students to transform basic sentences into more syntactically complex sentences, while preserving the original meaning (Limpo & Alves, 2013). For example, the sentences "The boy ate a sandwich" and "The boy was hungry" can be combined into "The boy ate a sandwich because he was hungry". Multiple studies, from the 1970's to the present, have shown beneficial effects of sentence-combining instruction on syntactic complexity and writing performance at all educational levels (see e.g., O'Hare, 1973; Strong, 1986; Hillocks, 1986; Saddler & Graham, 2005; Limpo & Alves, 2013; Saddler, Ellis-Robinson, & Asaro-Saddler, 2018). In these studies, the effects of sentence-combining instruction are typically compared to other types of instruction, such as more traditional grammar instruction (e.g., Saddler & Graham, 2005) or instruction focused on planning strategies (e.g., Limpo & Alves, 2013). However, experimental studies on the most effective way to teach sentence-combining are limited, which is the aim of the current study.

To the best of our knowledge, only Zimmerman and Kitsantas (2002) compared different types of sentence-combining instruction within one study. In their study, they investigated the effects of modeling type and social feedback on writing revision proficiency (measured through a sentence-combining task), self-satisfaction reactions, self-efficacy scores, and intrinsic interest in the task. More specifically, they compared observational learning from a coping model (a model gradually improving her performance over time), and from a mastery model (a model displaying a flawless execution of the task) to 'learning by doing', in which learners perform practice tasks themselves. Their findings showed that observational learning, from both the coping and mastery model, yielded higher performance scores than learning by doing: students who learned from the coping model outperformed students who learned from the mastery model who in turn outperformed those who learned without a model (no model condition) on all dependent measures. Social feedback, information given to a learner by others about the accuracy of his or her response, also positively affected these same measures, regardless of modeling condition.

In the current study, we conceptually replicate Zimmerman and Kitsantas (2002). We systematically compare observational learning from a coping model or mastery model with learning by doing (practicing). In addition, we add explicit reflection to all conditions in the design, since in observational learning reflective activities seem to be an integral part of the learning process (Braaksma, Van den Bergh, Rijlaarsdam, & Couzijn, 2001) while it is typically absent in learning by doing, which complicates the comparison between the methods. With this study, we aim to gain more insight into the effectiveness of instructional methods for teaching sentence-combining. In addition, by conceptually replicating Zimmerman and Kitsantas (2002), we can test the robustness of the theoretical implications of their study. This contributes to a growing body of literature on the effects of observational learning within the writing domain, since the sentence-combining task in the current study is of a less complex nature than the more extensive writing tasks (e.g., argumentative text, literature review) in most observational learning research within the writing domain.

Theoretical Background

Observational learning in writing research

In Zimmerman and Kitsantas' study (2002) students were taught sentence-combining through cognitive modeling, also referred to as observational learning. In observational learning, people learn from observing the behavior of models. Within the writing domain, it is very different from traditional learning to write methods in which students mostly learn by performing writing tasks, since students do not actually engage in actual write during observational learning (Braaksma et al., 2002). This arguably leads to a reduction in cognitive load. By observing and reflecting on the behavior of models, learners are supported in constructing a cognitive representation of that model's behavior and hopefully learn to connect this representation to their prior knowledge for future recall (Bandura, 1997; Bandura, 2016). In learning how to write, this means that learners gain insight into the writing processes and the emergence of the resulting text, by observing and reflecting on real models thinking aloud, rather than directly engaging in the cognitively demanding writing task itself, which allows learners to focus on learning instead of executing tasks (Couzijn, 1999; Rijlaarsdam & Couzijn, 2000).

In Zimmerman and Kitsantas, the writing task involved sentence-combining, which arguably is a rather basic written text. However, within the writing domain, the effects of observational learning have mostly been studied with different, more global types of writing tasks, often involving larger stretches of text, at different educational levels. For example, in a study by Braaksma et al. (2002), high school students wrote higher quality argumentative texts when they learned by observing models. With upper primary students, observational learning has been shown to lead to students applying more structured and goal-focused planning processes and to higher quality texts (Fidalgo, Torrance, Rijlaarsdam, van den Bergh, & Álvarez, 2015), even though Grenner et al. (2020) only found a short-term effect on text quality. Van de Weijer, Åkerlund, Johansson, & Sahlén (2019) did not find evidence for the effectiveness of observational learning in a small-scale study including university students with hearing impairment. Observational learning has also been reported to influence writing processes in poetry writing with high school students, even though this did not result in poems that were assessed as more creative (Groenendijk, Janssen, Rijlaarsdam, & Van den Bergh, 2013), and to positively affect writing revision skills in foreign language learners (Van Steendam et al., 2013).

Within the academic writing domain, the number of studies directly comparing observational learning and learning by doing is scarce, and the findings are inconclusive. Raedts, Rijlaarsdam, Van Waes and Daems (2007) found that students who learned from observation wrote literature reviews of higher quality (linking the source material more often and better organized structure) compared to the students who learned by performing writing exercises. However, in the studies reported on in Chapters 2 and 3, we found no differences in writing performance between students who learned to write an introduction of a scientific research paper by observing models and students who learned from writing exercises: the written academic texts were of equal quality. In addition, in a study by Callinan et al. (2018), observational learning was not more beneficial than post-submission feedback for essay writing skills of university students.

In most of these studies, students learned from watching different types of models: coping models, mastery models or a combination of both type of models. However, the effect of the type of model was not explicitly examined within these studies, with a few exceptions, most notably in Zimmerman & Kitsantas' study (2002). Previous research in different domains, such as the motor skill domain (see e.g., Shea, Wright, Wulf, & Whitacre, 2000; Wesch, Law, & Hall, 2007; Dempsey & Kaufmann, 2017) and more cognitive domains (Schunk & Hanson, 1985; 1989; Van Gog & Rummel, 2010), suggests, however, that the type of model can influence the learning process.

Type of Model in Observational Learning

In the learning sciences, the role of the type of model has been studied extensively. Some have advocated the use of mastery models, who exhibit a flawless performance of the task, since they display a standard of reference against which observers are able to detect their own errors and issue appropriate correction, which facilitates constructing a mental representation (Carroll & Bandura, 1987; Ferrari, 1999). Others have advocated the use of coping models, who gradually improve their performance, since they provide more information about strategy implementation and error correction than a mastery model does (Zimmerman & Kitsantas, 2002; Ferrari, 1999).

Only a few studies within writing research (Zimmerman & Kitsantas, 2002; Braaksma, Rijlaarsdam, & Van den Bergh, 2002) have investigated the effects of different types of models on writing performance. Zimmerman and Kitsantas (2002) investigated the effect of modeling type on writing revision proficiency, self-efficacy beliefs, self-reaction, and intrinsic interest with 72 undergraduate students. In their study, participants were first presented with a five-step solution strategy for rewriting multiple kernel sentences into a complete sentence and were told how to use this strategy on a sample problem. They then either watched a live coping or mastery model solving nine training problems, or they were not shown any video and instead studied the same nine training problems (no model condition). As a next step, there was a practice period for all student in which they were asked to solve twelve new training problems. As a final step, they were post-tested sequentially for self-efficacy, writing revision proficiency, attributions, intrinsic interest, and self-satisfaction. Zimmerman and Kitsantas (2002) found that students who had observed the coping model outperformed the students who had observed the mastery model in their ability to combine kernel sentences. Furthermore, they found that these students had more confidence in themselves (self-efficacy), were more satisfied with their results (self-reaction), showed more pleasure and intrinsic interest in the task than the students in the mastery model condition, who in turn scored better on all measures than the students in the no model condition.

Braaksma et al. (2002) examined the effects of model-similarity, that is, the perceived similarity in competence between model and observer on the effectiveness of observational learning in argumentative writing. Eighth-grade students learned how to write an argumentative text by either observing a weak and a strong model writing short argumentative text while thinking aloud, or by performing writing tasks themselves. In addition, the students in the observational learning conditions were asked to focus their reflection either on the strong model (observation/good-focus) or on the weak model (observation/weak-focus). The students who performed writing tasks were not asked to reflect on their own performance (control). Their findings suggest the effectiveness of the different types of models depends on the aptitude of the learners. From their study, however, we cannot draw conclusions on the effects of the weak versus the strong model, since the students in the modeling conditions were exposed to both, while in Zimmerman and Kitsantas (2002) the students were exposed to either the coping or mastery model. The operationalization of the two types of models also differs: the weak model in Braaksma et al. made mistakes and did not necessarily correct these, while the coping model in Zimmerman and Kitsantas improved her performance.

To the best of our knowledge, only these two studies have examined the type of model in writing research, and even though they both indicate that within the writing domain the type of model affects the learning process, the precise effects of the different types of models remain unclear. In addition, there might be a confounding factor in observational learning studies, namely reflection. In Braaksma et al. (2002), and most other studies on observational learning, (explicit) reflection is an integral part of observational learning. Learners are asked to monitor and evaluate the model's actions, which are precisely the activities assumed to play a central role in learning a cognitive skill, such as writing. This is however not the case for the learning by doing conditions in these previous studies. This lack of reflection in learning by doing complicates the comparison between both instructional methods.

Reflection in writing research

For most people, reflection is a natural and familiar process (Daudelin, 1996). By reflecting, learners explicitly attend to actions and performances and carefully process them (Wouters, Paas, & Van Merriënboer, 2009), which helps learners to develop new insights (Daudelin, 1996).

Within writing research, reflection is often incorporated in curricula aimed to promote self-regulated learning. According to Zimmerman (1989, p. 329), self-regulating students are students who are 'metacognitively, motivationally, and behaviorally active participants in their own learning process'. Three types of subprocesses are assumed to underlie self-regulated learning, namely self-observation, self-judgment, and self-reaction: in self-observation learners monitor their own actions, in self-judgment they evaluate their performance and by self-reactions learners respond to their performance outcomes. These characteristics could collectively be defined as reflection since they engage individuals in exploring their experiences in order to lead to new understandings and appreciations (Boud et al., 2005). Several studies on curricula promoting self-regulated learning to write, including one or more of these reflective activities, have reported positive relations between these activities and learning to write with different types of audiences (e.g., Fidalgo et al., 2015; MacArthur, Philippakos, & Ianetta, 2015; Graham et al., 2005; Graham, Kiuhara, & Harris, 2012; Zimmerman & Kitsantas, 2002, 1999).

Within the observational learning domain, reflection is, however, mostly not directed at the learner's own behavior, but rather directed at the actions and performances of models. Learners are for example asked to monitor the actions of models, which refers to activities in which deliberate attention is paid to some of one's behavior (Schunk, 1983). For example, learners are often asked to evaluate and elaborate on models' actions or products, by answering questions as 'Which model is doing better?', and 'Explain why the less good model is doing less well'. To our best knowledge, only a handful of studies have explicitly examined the role of reflection in observational learning within the writing domain (Braaksma et al., 2001; Callinan, Van der Zee, & Wilson, 2018; our study from Chapter 3).

Braaksma et al. (2001) investigated which learning activities were effective within observational learning. To do so, they examined the data (workbooks, pre- and post-tests) from an earlier study on observational learning by Couzijn (1995). In that study ninth-grade students were taught argumentative reading and writing by observational learning. Students observed models performing reading and writing tasks, and were explicitly asked to monitor, evaluate, and elaborate on the models' performance. The findings suggested that evaluating the performance of models and elaborating on the product of the models, contributed to argumentative writing skill. Braaksma and colleagues argued that students who evaluated and elaborated on the model's writing seemed to develop criteria for effective texts and writing processes which transferred to their own writing, yielding higher quality writing performance.

In a study with 77 first-year university students, Callinan et al. (2018) compared the effectiveness of post-submission feedback, observational learning without reflection and observational including reflection on developing essay writing skills. Students in the observational learning with reflection condition were encouraged to evaluate and discuss the peer model videos they watched, while the students in the observation only condition, merely watch the videos. Callinan et al. were not able to establish the relative contribution of pure observation and observation combined with reflection: they did not find differences between the three conditions.

A last study examining the role of reflection, is our study from Chapter 3, in which we focused on the effect of reflection on observational learning and learning by doing, in the context of an academic writing task. 111 undergraduate students were assigned to either an observational learning or learning by doing condition, with or without reflection. In the observational learning condition students learned by observing a weak and strong model's writing processes. In the learning by doing condition they learned by performing writing tasks. Half of the students reflected on either the models' or their own performance, by answering questions on the models' performance. In our study, neither reflection nor instructional method (observational learning and learning by doing) affected academic writing performance and self-efficacy beliefs. However, students did prefer reflecting over not reflecting, and learning by doing over observational learning.

From these studies, no clear view can yet be established on the effects of reflection on observational learning and learning by doing. To better understand how reflection affects observational learning and learning by doing, we systematically compare observational learning from a coping model, a mastery model, and learning by doing, with and without reflection in the current study.

Current study

In the current study, we conceptually replicate Zimmerman and Kitsantas (2002) in order to gain more insight into the effectiveness of instructional methods within the writing domain. Conceptual replication allows us to test the robustness of the theoretical implications (Zwaan, Etz, Lucas, & Donnellan, 2018). We chose this particular study because it is, to the best of our knowledge, the only study on sentence-combining in which the effectiveness of the instructional method is explicitly studied. In addition, the sentence-combining task allows us to investigate the role of the type of model and reflection with university students executing a less extensive writing task, since previous studies using more complex writing tasks (such as an academic literature review or introduction) yielded inconclusive results. Hence it seems important to first compare methods in a less complex writing task.

In the current study, we closely follow the procedures and measures of Zimmerman and Kitsantas (2002) which will be described in the method section (an extensive overview of the similarities and differences between the two studies is presented in Appendix A). However, we extend the original design in two aspects. First, we add reflection as a factor to the design. As mentioned, in most studies on observational learning, reflective activities are present in the observational learning conditions, but absent in the no model conditions, which complicates comparing the different conditions. By adding reflection, we attempt to make a more systematic comparison of observational learning (coping vs. mastery) and practice, with or without reflection. In order not to overcomplicate the design, we have taken out social feedback as a factor. Even though social feedback positively influenced writing performance, Zimmerman and Kitsantas report no interaction between modeling type and social feedback. We therefore do not expect our choice to influence possible effects of modeling type.

Second, we add two dependent measures. The first measure is implementation of the five-step solution strategy, to establish whether modeling type influences the execution of the performance. Braaksma et al. (2001) found that observational learning affects the orchestration of writing processes in argumentative writing, and we want to explore whether this transposes to sentence-combining tasks. The second measure is satisfaction with the instructional method. Often there is a mismatch between the way that someone prefers to learn and what actually leads to effective and efficient learning (Kirschner, 2017). An early, illustrative discussion of this can be found in a review by Clark (1982) who found that students often report enjoying the method from which they learn the least. In Chapter 3, we found that students preferred learning by doing over observational learning, and reflecting over not reflecting, even though this was not related to performance and self-efficacy. We aim to explore this effect further in the current study. We formulated three research questions:

Research Question 1: How does observational learning from a coping model or a mastery model influence writing revision proficiency, self-efficacy beliefs, self-reaction, satisfaction with the instructional method and implementation of the solution strategy, compared to learning by doing?

Since our study is a conceptual replication, our hypotheses are similar to Zimmerman and Kitsantas (2002). We hypothesize that modeling type influences writing revision proficiency. More specifically, we expect that observing a model leads to higher scores on all dependent measures than not observing a model (H1a) and in addition, that observing a coping model leads to higher scores than observing a mastery model (H1b).

Research Question 2: How does reflection affect writing revision proficiency, self-efficacy beliefs, self-reaction, satisfaction with the instructional method and implementation of the solution strategy?

Because of the inconclusive nature of earlier findings on reflection (Braaksma et al., 2001; our study from Chapter 3), we formulate no specific hypothesis here.

Research Question 3: What is the interplay between observational learning from a coping model, a mastery model or no model, and reflection?

Based on earlier work, we hypothesize an interaction between modeling type and reflection, in which we expect reflection to reduce the effect of modeling type on all measures (H2).

Method

Design

A 3 x 2 between subjects factorial design was used in this study, with modeling type (no model, coping model, mastery model) and reflection (present, absent) as factors. We measured the effects on five dependent variables: writing revision proficiency, self-efficacy, self-reaction, satisfaction with the instructional activities and the implementation of a solution strategy. Table 1 displays an overview of the different stages and corresponding tasks within the study. The timing of the study was determined after a pilot study with twelve participants.

Table 1.
Set-up of the study

	Stage	Task	Time in Minutes
0	Pretest	Filling out consent form Grammar test	5
1	Introduction	Explanation of the task by experiment leader Introduction five-step solution strategy by experiment leader Example writing revision problem Pre-measurement self-efficacy beliefs	6
2	Intervention (modeling type)	No model conditions: solving practice problems Coping and mastery model conditions: watching a video of a model solving practice problems	10
3	*Intervention* (reflection)	Only for the reflection conditions: answering reflection questions	5
4	Posttest 1	Measurement writing revision proficiency, self-efficacy, self--reaction and the use of the five-step solution strategy	13
5	Practice Round	Solving practice problems	10
6	Posttest 2	Measurement writing revision proficiency, self-efficacy, self-reaction	13

Participants

The participants for his study were recruited from a participants' pool of the departments Communication and Information Sciences, and Psychology of Tilburg University in the Netherlands. A total of 108 undergraduate students (69 women, 39 men) between the ages of 17 and 27 (M = 21,4 years, SD = 1.83) participated in the study for partial course credits. All participants were Dutch native speakers. See Table 2 for the distribution of participants over conditions.

Table 2.
Number of Participants per Condition

	No Model	Coping Model	Mastery Model	Total
No Reflection	17	19	19	55
Reflection	18	17	18	53
Total	35	36	37	108

Materials

Writing Revision Problems. In line with Zimmerman and Kitsantas (2002), the writing tasks in this study were writing revision problems in which three to seven kernel sentences had to be combined into one non-repetitive sentence. These sentence combining problems were, as in Zimmerman and Kitsantas, selected from Strong's (1981) Sentence Combining: A Composing Book and subsequently translated into Dutch. A total of 14 problems was included in the study: one problem was used as an example in the introduction of the study. Four problems were used in the modeling videos. These same four problems were used as practice problems in the no-model-conditions. Posttest 1 consisted of three problems. In the practice round that followed posttest 1 another three problems were used. Posttest 2 again consisted of three problems. An example of a problem can be found in Figure 1. Please note that in the current study, all writing revision problems were presented in Dutch.

Figure 1.
Example of a writing revision problem.

Kernel sentences:	*An optimally combined sentence*:
The aqueduct is located in Segovia in Spain. It was built by the Romans. It was built around AD 100 It was built under the rule of Emperor Trajan.	The aqueduct, which is located in Segovia in Spain, was built around AD 100 by the Romans under the rule of Emperor Trajan.

Five-step solution strategy. All participants were presented with a five-step solution strategy for the writing revisions problem during the introduction of the experiment. These five steps were taken from Zimmerman and Kitsantas (2002) and translated into Dutch. The five steps were: (a) place in bold all the words standing for new concepts in each sentence, (b) strike-through the words that refer to the same concept, (c) combine all the bolded words into written clauses about the same concept, (d) number the clauses in order of importance, and (e) build the final sentence around the most important clauses and insert less important clauses where they belong using connecting words.

Videos in observational learning (coping and mastery model). Two different videos were recorded, one of a coping model solving four writing revision problems and one of a mastery model solving the same problems. We chose video modeling instead of live modeling (which was used in Zimmerman and Kitsantas, 2002), to ensure that all participants within conditions received identical treatments.

The same female model appeared in both videos. The model was a student actor who received a script for each problem and was instructed to think aloud while solving the problems. The script contained concrete suggestions for remarks to be made, but the model was also allowed to give spontaneous input in order to make the videos as natural as possible.

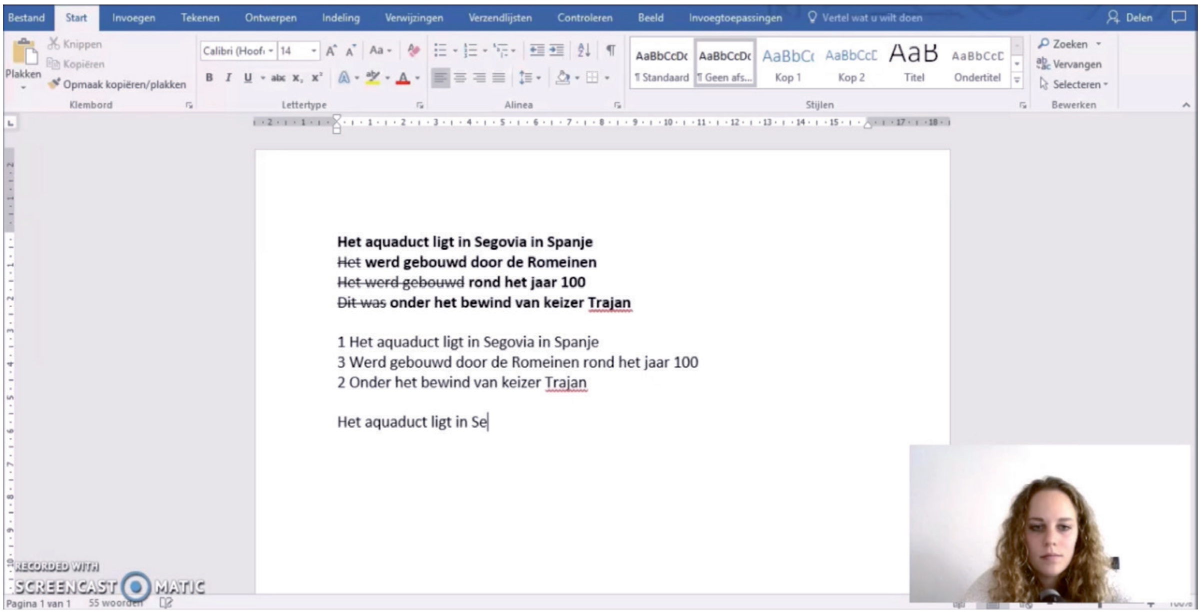
In the coping model video, the model started out by making several errors in the application of the five-step solution strategy. In the first problem she made several mistakes: she, for example, wrongly placed a repeated concept in bold. In solving the second and third problem she made one mistake: in the second problem the model forgot to strike-through a repeated concept, and in the third problem, she accidentally forgot to place a new concept in bold. The model noticed all her mistakes and corrected them later on, leading to a correctly combined sentence in the end. The fourth problem was solved flawlessly by the model.

In the mastery model video, the model started out with a perfect application of the five-step solution strategy and maintained this throughout the video.

The model commented in both videos on her own performance, by describing the solution strategies she was using. The solutions for each problem the model presented, a combined nonrepetitive sentence, were identical in the coping and mastery model video.

The videos were recorded using Screencast-o-matic (https://screencast-o-matic.com) which allows for simultaneous, picture-in-picture recording. Each video contained a recording of the model working on the computer, the model's voice and the computer screen the model was working on in Word. Figure 2 displays a still of the model solving the second practice problem in the coping model condition. Both videos lasted between 7 to 8 minutes.

Figure 2.
Still of the model solving a practice problem.



Reflection Questions. To prompt reflection, participants in the reflection conditions were asked to answer seven questions. In the coping and mastery model conditions the questions were the following: (1) What kind of words are crossed out by the model and what kind of words are printed in bold? And what is the objective of this? (2) Why do you think it is important that the model divides the information into clauses? (3) Why do you think the model puts the clauses in order of importance? (4) Why is adding connecting words important for this task? (5) What can the model do better next time? (6) In what order did the model tackle the task? (7) Were there any steps missing from the five-step solution strategy? If so, which ones?

In the no model condition, participants had to reflect on their own behavior, instead of on a model. Therefore, we adapted the reflection questions from the coping and mastery model conditions, in order to make them suitable for self-reflection. For example, we rephrased question (2) as "Why do you think it is important that you divide the information into clauses?" and question (5) as "What can you do better next time?"

In the reflection questions all five steps from the solution strategy were represented which allowed participants to reflect on and process the different elements of the solution strategy and their relevance for solving the problem in all conditions.

Measures

Grammar. Prior to the interventions, participants filled out a grammar test in order to check for possible initial differences in grammatical ability. The test was developed by the Language Center of the university in which the experiment was executed and has been used for over a decade as a diagnostic instrument for undergraduate students at this university. Grammar was tested with 25 items, containing congruency problems (8), verb conjugations (5) and endophoric expressions (12). Per item one point could be scored, resulting in a possible minimum score of zero and a maximum score of 25 points.

Writing revision proficiency. Writing revision proficiency was measured in two posttests, each containing three revision problems. The participants were asked to create a single non-repetitive sentence for each revision problem. The answers were scored identical to Zimmerman and Kitsantas (2002). That means that for each sentence there was a maximum score of 100 points. If students wrongfully left out a word or if they repeated a word, 1 point was deducted. For failing to combine the sentence in a grammatically correct manner, 40 points were deducted. 10 points were deducted for each misplaced clause. If the student did not follow the instructions, for example by creating two sentences or creating a revision sentence that included only a few kernel sentences, 40 points were deducted, as well as 1 point for each missing word. We calculated the final score for each posttest as the average score for the three problems per posttest. Twenty percent of the solutions were rescored by a second coder (inter-rater reliability Pretest 1, Pearson's $r = .91$, inter-rater reliability Posttest 1, Pearson's $r = .91$; inter-rater reliability Posttest 2, Pearson's $r = .80$).

Self-efficacy. The self-efficacy measure was identical to the one used by Zimmerman and Kitsantas (2002), which was based on Bandura and Schunk (1981). Participants were asked how sure they were they would be able to solve a similar sentence-combining problem. They could answer the questions by placing a number from 10 to 100 in the space provided on their hand-out. Written descriptions were provided for the following scores: 10 (not sure), 40 (somewhat sure), 70 (pretty sure), and 100 (very sure). Self-efficacy was measured three times during the experiment: once in the introduction phase directly after the example problem was discussed, once following the modeling and reflection interventions and finally directly following the practice round.

Self-reaction. Self-reaction was also measured identical to Zimmerman and Kitsantas (2002). Participants were asked to answer how satisfied they were with the sentence they constructed, by using a scale ranging from 10 to 100. Written descriptions were provided for the following points: 10 (not satisfied), 40 (somewhat satisfied), 70 (pretty satisfied), and 100 (very satisfied). Self-reaction was measured twice: once after completing posttest 1 and once after completing posttest 2.

Satisfaction. After posttest 2, participants were asked to evaluate how satisfied they were with the modeling intervention. Participants had to indicate on a scale from 1 to 7 (1 = not at all, 7 = very much) to what extent they agreed with four items. Examples of items are: "The video / practice problems helped me in completing the test" and "I learned a lot from the video / practice problems". The average score of the four items (Cronbach's alpha = .84) was used as the participants' satisfaction score.

Implementation of the five-step solution strategy. After posttest 1, we measured how the participants implemented the five-step solution strategy. We did this by analyzing the screen recordings of posttest 1, by looking at three aspects: (1) execution of the steps, (2) number of skipped steps, and (3) order of the steps.

(1) Execution of the steps. We counted in how many of the three problems in posttest 1, they skipped one or more steps before starting to construct the combined sentence. For example, in Figure 3 an example is displayed in which a participant skipped step b: the participant did combine bolded concepts, but without striking through words that refer to the same concept first.

Figure 3.
Example of a participant who skipped step b.

If they used all steps in all three writing revision problems, they received a score of 0. If they skipped one or more steps in one problem, they received a score of 1, in two problems a score of 2, and in all problems a score of 3.

(2) *Number of skipped steps in all three problems*. We counted the total number of steps skipped for all three problems. Since it was possible for participants to skip more than one step per problem, we counted this separately from (1). The lowest possible score was 0 (using all steps in all three problems) and the highest 18 (skipping all steps in all three problems).

(3) *Order of the steps*. We looked at whether the participants followed the prescribed order (a to e) of the five-step solution strategy in each of the problems. If they used the prescribed order in all three problems, they received a score of 0, in two problems a score of 1, in one problem a score of 2, and if they used a different order in all three problems, they received a score of 3.

Procedure

The experiment was organized over different time slots. Each time slot was assigned to one of six conditions. Participants could enroll themselves for a time slot, unaware of the condition of that particular time slot. Per time slot two to six participants enrolled.

The experiment took place in a regular classroom, with a beamer and for each participant an available laptop. The laptops were positioned in such a way that the participants could not look at each other's screens. Next to the laptops, there was a pen for the participants, a consent form, a hand-out with the five-step solution strategy and a hand-out with the self-efficacy measures, the reflection questions (only for the reflection conditions), and the self-reaction and satisfaction questions. The two post-tests were administered on the laptop.

Two experiment leaders were present during the session. The participants were welcomed by one of experiment leaders who used a script so that every participant per condition received the same instructions. First, the participants were asked to complete the grammar test on the laptop. They were also asked to answer demographical questions, such as sex, age, and educational level. Then, the actual experiment started which consisted of either five (no reflection conditions) or six stages (see Table 1).

The participants were asked to read the five-step solution strategy on their hand-out. After this, one of the experiment leaders showed the example writing revision problem and how this problem could be solved using the solution strategy. This was presented via the beamer, accompanied by a brief explanation by the experiment leader. Then the participants' self-efficacy was measured for the first time. This was followed by the modeling and reflection intervention. In the no model-no reflection condition, participants were asked to practice with four practice problems for ten minutes. This was the same for the no-model-with reflection condition. This last group, however, answered the seven reflection questions after practicing. In the coping model and mastery model conditions without reflection, the participants watched the modeling video, which was presented via de beamer. This was the same for the coping and mastery model conditions with reflection. In these conditions watching the video was, however, followed by answering the reflection questions.

After the modeling and reflection intervention, the procedure was identical for all conditions. The participants were asked to complete posttest 1, which consisted of the second measurement of self-efficacy, followed by three writing revision problems that had to be completed in a Word Online document, and the self-reaction questions in the hand-out. The participants' screens were recorded with Screencast-o-matic during posttest 1 in order for us to be able to analyze the implementation of the five-step solution strategy. After 13 minutes, the experiment leader stopped the screen recording and the participants' documents were automatically saved by Word Online. A 10-minute practice round followed, in which participants were asked to try and solve three additional practice problems. Then the participants completed posttest 2, following the same procedures as posttest 1. However, their screens were not recorded this time, since we were mostly interested in the direct effects of modeling type on the strategy implementation, and the questions measuring participants' satisfaction with the modeling interventions were added.

Statistical Analyses

We conducted a one-way ANOVA to test for grammatical ability. For writing revision proficiency, self-efficacy and self-reaction, we used a repeated measure ANOVA. Mauchly's test for sphericity was used to test for homogeneity of variance, and when this test was significant, we applied a Greenhouse-Geisser correction on the degrees of freedom. We used a one-way ANOVA for measuring the satisfaction with the interventions and a MANOVA to analyze the implementation of the solution strategy. For the MANOVA Levene's test was used to test the homogeneity of variance. When this test was violated, we used Pillai's trace. For all analyses, Instructional Method and Reflection were included as factors.

Results

Grammatical Ability

Prior to the experiment, we measured the participants' grammatical ability. No significant differences were found between the modeling groups ($F(2, 102) = 0.88$, $p = .42$) and the reflection groups ($F(2, 102) = 2.78$, $p = .10$). This implies that there is no reason to assume there were a priori differences between conditions in grammatical ability. Table 3 presents an overview of the mean scores per condition.

Table 3.
Mean Grammar Scores (SDs) per Condition (Modeling Type and Reflection)

	No Model	Coping Model	Mastery Model
No Reflection	12.65 (2.34)	11.53 (2.04)	12.21 (3.47)
Reflection	13.06 (1.73)	12.71 (2.05)	12.89 (1.97)

Writing Revision Proficiency

The repeated measures ANOVA revealed no main within-subjects effect of testing moment, *F*(1, 102) = 1.77, *p* = .19): there were no differences between the scores in posttest 1 (*M* = 74.57, 95% CI [70.99, 78.14]) and in posttest 2, after participants had an extra practice period (*M* = 76.95, 95% CI [74.18, 79.72]).

No significant main effect of modeling type on writing revision proficiency was found, *F*(2, 102) = 0.73, *p* = .49). Participants who learned without a model (*M* = 77.01, 95% CI [72.34, 78.14]) scored equally well as those who learned with a coping (*M* = 73.47, 95% CI [68.85, 78.08]) or mastery model (*M* = 76.79, 95% CI [72.25, 81.34]).

Reflection had no main effect on writing revision proficiency either, *F*(1, 102) = 0.99, *p* = .32. The writing revision scores in the conditions without reflection (*M* = 74.42, 95% CI [70.69, 78.15]) did not differ from those in conditions with reflection (*M* = 77.09, 95% CI [73.29, 80.89]).

In addition, none of the interaction effects were found to be significant (F_s < 1.96, p_s > .15). An overview of the mean scores for writing revision proficiency in all conditions is presented in Table 4.

Table 4.
Writing Revision Proficiency (SDs) as a Function of Modeling Type, Reflection and Testing Moment

	No Model		Coping		Mastery	
Posttest nr	1	2	1	2	1	2
No reflection	74.27 (21.69)	74.14 (14.52)	66.44 (19.04)	73.51 (13.33)	79.98 (17.96)	78.19 (16.85)
Reflection	78.59 (18.80)	81.04 (12.56)	73.31 (18.94)	80.61 (14.75)	74.80 (15.60)	74.20 (14.56)

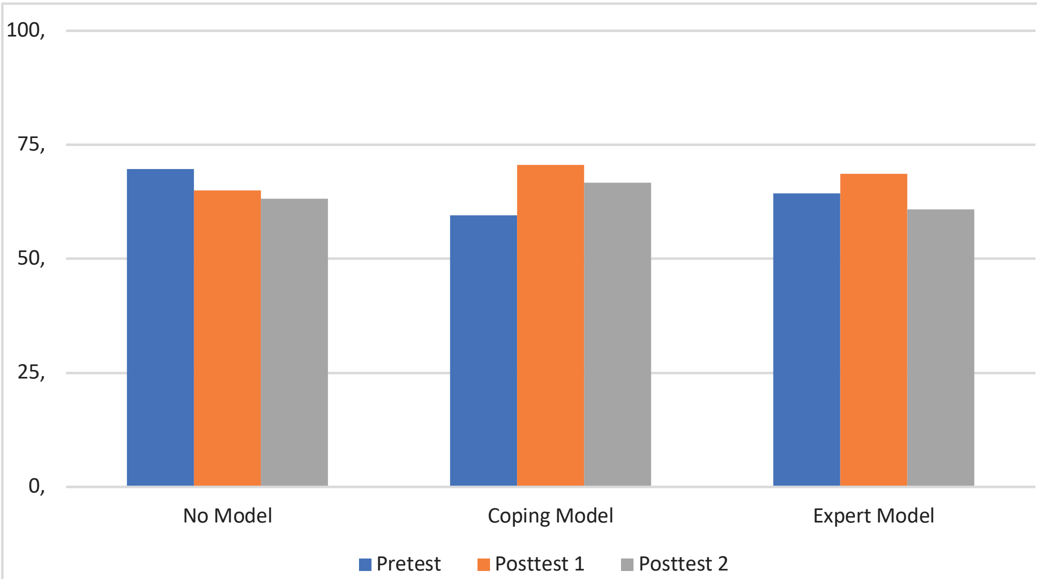
Self-efficacy

A repeated measures ANOVA was conducted in order to check for differences between the three measurements of self-efficacy (directly following an example problem, directly following the modeling and reflection intervention in posttest 1 and directly following the practice period in posttest 2). We applied a Greenhouse-Geisser correction on the degrees of freedoms, since Mauchly's test for sphericity was significant. There was a main within-subjects effect of testing moment on self-efficacy, *F*(1.81, 181.30)= 5.19, *p* <.01, η_p^2 = .05.

There was a significant increase between the pretest of self-efficacy and the self-efficacy score in posttest 1 (*M-diff* = -3.55, $p = .02$, 95% CI = [-6.69, -.42]) and a significant decrease in self-efficacy between posttest 1 and 2 (*M-diff* = -4.51, $p < .01$, 95% CI = [1.05, 7.98]). The mean difference between the pretest and posttest 2 (*M-diff* = .96) turned out not to be significant, $p = 1.00$, 95% CI = [-3.15, 5.07].

We found a significant interaction between testing moment and modeling type, $F(3.63, 181.30) = 6.13$, $p < .001$, $\eta_p^2 = .11$. Figure 4 illustrates this interaction. While the self-efficacy scores drop in the no modeling condition in both posttest 1 and 2, the self-efficacy scores in both the modeling conditions first increase in posttest 1 and then show a decrease in posttest 2.

Figure 4.
Self-efficacy as a function of Modeling Type and Testing Moment



The other interactions turned out not to be significant $F_s < 6.13$, $p_s > .47$*)*. Table 5 provides an overview of the mean scores for self-efficacy in all conditions.

Table 5.
Self-efficacy (SDs) as a Function of Modeling Type, Reflection and Testing Moment

	No Model			Coping			Mastery		
Test nr	Pretest	PT1	PT2	Pretest	PT1	PT2	Pretest	PT1	PT2
No reflection	74.69 (16.78)	68.75 (13.23)	65.50 (16.43)	55.79 (20.57)	69.74 (15.50)	66.32 (20.54)	64.21 (17.89)	68.42 (18.79)	58.95 (20.32)
Reflection	64.71 (13.97)	61.29 (18.72)	60.88 (23.67)	63.41 (18.89)	71.47 (14.66)	67.06 (18.88)	64.44 (17.96)	68.89 (15.10)	62.78 (16.65)

Note: PT refers to Posttest

Self-reaction

A significant main effect of testing moment on self-reaction scores was found, *F*(1, 101) = 23.65, $p < .001$, $\eta p^2 = .19$. In posttest 2 (*M* = 65.10, 95% CI [61.37, 68.83] the scores increased significantly compared to posttest 1 (*M* = 56.95, 95% CI [53.11, 60.79]). The mean difference between trial 1 and 2 (*M-diff* = -8.15) was significant, $p < .001$, 95% CI = [-11.47, -4.82]. This indicates that participants were more satisfied with their solutions in the second posttest after the practice period than in the first posttest directly following the modeling and reflection intervention.

No significant main effect of modeling type, (F (2, 101) = 1.67, *p* = .19) or reflection (F (1, 101) = 0.07, *p* = .79) was found, indicating that no model (*M* = 62.60, 95% CI [56.57, 68.63]), observing a coping model (*M* = 63.77, CI [57.91, 69.63]) and observing a mastery model (*M* = 56.70, 95% CI [50.93, 62.48]) do not lead to significantly different self-reaction scores, and neither does not reflecting (*M* = 60.57, 95% CI [55.78 , 65.36]) or reflecting (*M* = 61.48, CI [56.66 , 66.30]). In addition, no significant interaction effects were found, $F_s < 1.54$, $p_s > .22$. See Table 6 for an overview of the self-reaction scores.

Table 6.
Self-reaction (SDs) as a Function of Modeling Type, Reflection and Testing Moment

	No Model		Coping		Mastery	
	PT1	PT2	PT1	PT2	PT1	PT2
No reflection	59.06 (17.53)	70.56 (17.20)	59.79 (19.02)	66.63 (22.68)	48.42 (23.34)	58.95 (18.68)
Reflection	56.39 (21.68)	64.39 (24.85)	59.71 (20.35)	68.94 (15.66)	58.33 (16.72)	61.11 (14.71)

Evaluation of the Instructional Method

The instructional activities (modeling type: coping, mastery or no model, and reflection: absent or present) were evaluated by asking participants how satisfied they were with the activities. The ANOVA showed a significant main effect of modeling type, *F*(1, 99) = 4.04, *p* = .021, $\eta_p^2 = .07$. Pairwise comparisons revealed that the participants who learned without a model were the least satisfied (*M* = 4.72, 95% CI [4.37, 5.06]), followed by participants who learned from a coping (*M* = 5.30, 95% CI [4.96, 5.65]), and mastery model (*M* = 5.34, 95% CI [5.00, 5.67]). The mean difference between the no model and mastery model condition turned out to be significant (*M-diff* = -.62, $p < .05$, 95% CI = [-1.21, -.03]). In addition, there was no effect of reflection, *F*(1, 99) = .035, *p* = .85, and no significant interaction, *F*(1, 99) = .18, *p* = .84. Table 7 displays the mean satisfaction scores per condition.

Table 7.
Satisfaction (SDs) with the Instructional Method as Function of Modeling Type and Reflection

	No Model	Coping Model	Mastery Model
No Reflection	4.66 (0.98)	5.39 (1.03)	5.37 (1.12)
Reflection	4.78 (1.16)	5.22 (0.80)	5.31 (0.96)

Explorative Analysis Implementation of the Five-step-solution Strategy

We examined how the participants implemented the solution strategy by performing a MANOVA with modeling type and reflection as factors and number of tasks in which a step was skipped, total number of skipped steps and number of tasks in which the prescribed order was followed as dependent variables. When the assumption of homogeneity of variance was violated, we used Pillai's trace.

The multivariate result was significant for modeling type, Pillai's Trace = .31, $F(6, 184) = 5.63$, $p < .001$, $\eta_p^2 = .16$, indicating a difference in the implementation of the solution strategy between the no model, coping model and mastery model conditions. We found no effect for reflection, Pillai's Trace = .00, $F(3, 91) = 0.10$, $p = .96$.

A significant main effect for modeling type on total number of skipped steps was found, $F(2, 93) = 16.62$, $p < .001$, $\eta_p^2 = .26$. Participants in the no model conditions ($M = 3.07$, 95% CI [2.41, 3.73]) skipped more steps than participants in the coping model conditions ($M = 0.73$, 95% CI [0.05, 1.40]) who in turn skipped more steps than those in the mastery model conditions ($M = 0.33$, 95% CI [-0.33, 0.99]). The mean difference between the no model and coping model conditions turned out be significant (*M-diff* = 2.34, $p < .001$, 95% CI = 1.18, 3.50]), and so did the mean difference between the no model and mastery conditions (*M-diff* = 2.73, $p < .001$, 95% CI = 1.18, 3.50]). However, the mean difference between the coping and mastery model conditions was not significant (*M-diff* = 0.39, $p = 1.00$, 95% CI = -0.77, 1.55]). For reflection, there was no effect on the number of skipped steps, $F(1, 93) = .28$, $p = .60$.

We found no effects for modeling type and reflection on the number of tasks in which steps were skipped and for the number of tasks in which the prescribed order was followed, all $F_s < 1.39$, all $p_s > .23$). None of the interactions were significant either (all $F_s < 1.59$, all $p_s > .051$). In Table 8 the mean scores of the implementation measures are presented.

Table 8.
Implementation Solution Strategy Scores (SDs) as a Function of Modeling Type and Reflection

	Experimental Group											
	No Reflection						Reflection					
	No Model		Coping		Mastery		No Model		Coping		Mastery	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Tasks with skipped steps	0.75	1.00	0.44	0.63	0.17	0.51	0.28	0.57	0.73	0.70	0.44	0.83
Skipped steps	2.75	2.75	0.63	0.89	0.22	0.73	2.83	2.96	0.73	0.88	0.63	1.26
Order of the steps	0.56	0.81	0.69	0.70	0.39	0.61	0.33	0.59	0.73	0.88	0.50	0.63

Correlations

To further explore the relationships between writing revision proficiency, self-efficacy and self-reaction, we used a Pearson's correlation test. In the first posttest, the results show a weak but significant correlation between writing revision proficiency and self-reaction, $r = .30$, $p < .01$. There was a moderate correlation between self-efficacy and self-reaction, $r = .42$, $p < .001$. The correlation between writing revisions proficiency and self-efficacy was, however, very weak and non-significant, $r = .08$, $p = .42$. See Table 9 for an overview.

Table 9.
Correlations between Dependent Variables in Posttest 1

	Writing Revision	Self-efficacy	Self-reaction
Writing Revision	-		
Self-efficacy	.08	-	
Self-reaction	.30**	.42**	-

** significant at the 0.01 level (2-tailed)

For the second posttest, there is a strong correlation between self-efficacy and self-reaction, $r = .76$, $p < .001$. The correlation between writing proficiency and self-efficacy is weak but significant, $r = .22$, $p < .05$. The correlation between writing revision proficiency and self-reaction is however weak and non-significant, $r = .22$, $p = .08$. In Table 10 an overview can be found.

Table 10.
Correlations between Dependent Variables in Posttest 2

	Writing Revision	Self-efficacy	Self-reaction
Writing Revision	-		
Self-efficacy	.22*	-	
Self-reaction	.17	.76**	-

** significant at the 0.01 level (2-tailed), * significant at the 0.05 level (2-tailed)

In Table 11 the correlations between the dependent measures in posttest 1 and 2 are displayed. Inspection of the table reveals that there were significant correlations between self-reaction and self-efficacy. Self-reaction in posttest 1 was positively correlated to self-efficacy in posttest 2 (r = .80, p < .001) and vice versa (r = .54, p < .001). There were also strong correlations between self-efficacy in posttest 1 and 2 (r = .68, p < .001), and between self-reaction in posttest 1 and 2 (r = .62, p < .001).

There was only a moderate correlation between writing proficiency in posttest 1 and posttest 2 (r = .41, p < .001) and weak correlations between self-reaction in posttest 1 and writing proficiency in posttest 2 (r = .23, p < .05), and writing proficiency in posttest 1 and self-efficacy in posttest 2 (r = .27, p < .01).

Table 11
Correlations between the Dependent Variables in Posttest 1 and Posttest 2

	Writing Proficiency 1	Self-efficacy 1	Self-reaction 1
Writing Proficiency 2	.41**	.093	.23*
Self-efficacy 2	.27**	.68**	.80**
Self-reaction 2	.08	.54**	.62**

** significant at the 0.01 level (2-tailed), * significant at the 0.05 level (2-tailed)

General Discussion

In the current study, we systematically compared instructional methods for learning how to combine sentences. More specifically, we investigated the effects of learning from different types of models (learning by observation from a mastery or coping model) compared to learning without a model (learning by doing) on writing revision proficiency, self-efficacy beliefs, self-reaction, satisfaction with the instructional method and the implementation of a solution strategy by conceptually replicating Zimmerman and Kitsantas (2002). In addition, we studied the role of reflection in learning to write.

Modeling type: coping model and mastery model versus learning by doing

With our first research question we aimed to investigate to what extent observational learning from a coping model or a mastery model influences writing revision proficiency, self-efficacy beliefs, self-reaction, satisfaction with the instructional method and implementation of the solution strategy, compared to learning by doing.

We found that modeling type did not affect writing revision proficiency and self-reaction. In fact, writing revision proficiency did not improve at all between the first and the second posttest, which was the case for all modeling conditions. The self-reaction scores did increase, also irrespective of the type of model: students were more satisfied with their answers after a practice round than directly following the modeling intervention, even though their writing proficiency scores did not improve. These findings imply that a practice round does not necessarily lead to more learning. It should be noted that we, in line with Zimmerman and Kitsantas (2002), did not administer a writing revision pretest since it could be argued that a pretest interacts with students' responses to the treatments. We therefore used a random assignment in order to detect possible initial differences between conditions, as did Zimmerman and Kitsantas. This means, however, that we cannot firmly establish the students' initial writing revision proficiency or how much the students learned from the modeling intervention alone.

For self-efficacy, we found an interaction between testing moment and modeling type. After observing a coping or mastery model, students felt more confident in their ability to successfully complete the sentence-combining tasks, while students who merely practiced felt less confident, which corresponds with earlier studies reporting positive effects of learning from models on self-efficacy (see e.g., Schunk & Zimmerman, 2007; Law & Hall, 2009). However, in our study, the self-efficacy scores decreased for all students after the practice round preceding posttest 2, implying that practicing lead to less confidence. Also noteworthy is that, in line with Zimmerman and Kitsantas (2002), self-efficacy was, in both posttests, not related to writing proficiency. This seems to indicate that students are not capable of assessing their own level of learning, which has also been reported in previous studies (see e.g., Pajaras & Graham, 1999). We did find a strong correlation between self-reaction after posttest 1 and self-efficacy before posttest 2: students who were more satisfied with their answers after the first posttest were also more confident in their ability to successfully combine sentences before the second posttest.

Interestingly, even though no differences in writing revision proficiency were found between conditions, our findings indicate that students still preferred learning from observing a mastery model, over observing a coping model and learning by doing. Students who observed a mastery model found the intervention more useful and helpful than the students in the other conditions. This in line with Kirschner (2017) who noted that the type of learning that students prefer is not necessarily the most effective or efficient way. This result might be explained by the fact that learners perceive experts as more professional and trustworthy than for example peers (Yang et al., 2006). In addition, we also found that students in the coping and mastery model conditions seemed to follow the five-step-solution strategy more closely.

In sum, our data do not support H1a in which we expected that students who learned from observing a coping model would outperform students who learned from a mastery model. Only for satisfaction with the instructional method did we find a difference between the mastery and coping model, but contrary to what was expected, the mastery model was found to be *more* useful than the coping model. On the other measures we did not find differences between the coping and mastery model conditions. Hypothesis H1b in which we hypothesized that the mastery and coping model conditions would outperform the no model condition was partially supported. There were no differences between learning by doing and learning from the mastery and coping model with regards to writing revision proficiency and self-reaction. Self-efficacy did differ between modeling conditions: self-efficacy dropped for the students who learned by doing while it initially increased for the coping and mastery model conditions and only then dropped. We also found that students who learned by doing implemented the solution strategy less accurately than the students who learned from a coping or mastery model, but this did not lead to higher writing revision proficiency.

Reflection and The Interplay between Reflection and Modeling Type

Our second and third research question dealt with the role of reflection in learning to write, with and without a model. We investigated the effect of reflection on writing revision proficiency, self-efficacy beliefs, self-reaction, satisfaction with the instructional method and the implementation of a solution strategy. We expected reflection to mitigate the effects of modeling type (H2), but our findings did not support this. Students in the reflection and no reflection conditions scored equally high on all dependent measures. This is in line with what we found in Chapter 2. In addition, for none of the measures we found an interaction between modeling type and reflection. This indicates that reflection in our study did not affect learning to write.

On Replicating Zimmerman and Kitsantas

With our study we were not able to replicate the findings of Zimmerman and Kitsantas (2002). Students who learned from observing models seemed to more closely follow a sentence combining solution strategy, but this did not lead to a higher writing revision proficiency.

How can we explain these results? One might argue that the replication was not executed accurately. However, our materials and procedure closely followed those of Zimmerman and Kitsantas. In addition, it is unlikely that deviations from the original design would explain the different findings, since these were either not expected to influence the main effects (such as eliminating social feedback and adding reflection as a factor, as argued in the method section) or the deviations could be considered improvements (for example the use of video model instead of a live model, which guaranteed all participants within a condition to receive an identical treatment).

One possible explanation lies in the type of task in our study: sentence combining. In order to stay as close as possible to the original materials and procedures, we have used tasks from *Sentence Combining: a composing book* by Strong (1981) as did Zimmerman and Kitsantas and translated these into Dutch. The six tasks from the posttests all contained seven kernel sentences and it was possible to formulate one non-repetitive sentence from those kernel sentences. However, the complexity of combining kernel sentences may differ between English and Dutch. For example, Dutch is a subject–object–verb (SOV) language. In main clauses, the finite verb follows the first constituent because of a verb-second rule; this results in SVO orders. In subclauses on the other hand, the order is SOV (Koster, 1974). In addition, some connectives (*en (and), maar (but), want (because), of (or)*) are followed by a verb-second main clause while other connectives (e.g., *omdat (because), terwijl (while), nadat (after)*) are followed by a SOV subclause. This might have overcomplicated the tasks for students. In comparison, the mean scores in all six conditions in our study vary, on a scale between 0 to 100, between 70 and 80 points, while in Zimmerman and Kitsantas the scores vary between 70 and 96. However, even though there may be subtle differences between Dutch and English, it seems unlikely that this is the main explanation of why we did not replicate Zimmerman and Kitsantas, considering the substantial similarities also present between these two Germanic languages.

Another explanation might be the five-step-solution strategy. In the observational learning videos the focus was on this strategy: the model showed the use of the different steps in order to combine the kernel sentences, and the reflection questions also directed attention to these steps. However, we cannot be sure to what extent following the solution strategy was actually supportive in the ability to combine sentences effectively. In our study, students who followed the solution strategy more closely did not outperform the others. This could indicate that solution strategy was not as effective as expected. It is possible that other factors (e.g., knowledge of main and subclauses in Dutch) might have influenced the results, even though we did not find a priori differences between conditions on a grammar test. In addition, even though the students were able to see the emerging sentence, the quality of the combined sentence was not the focus of the videos or reflection. This might have caused students to have an unclear view on an optimally combined sentence.

It is worth noting, however, that in the two studies of chapter 2 and 3, concerning observational learning of a more complex writing task, i.e., the introduction of a research paper, we also found no differences between students who learned from observing models or from executing writing tasks themselves. In Callinan et al.'s study (2018), again with university students learning to write a complex text, also no effects of observational learning were reported when compared to post-submission feedback. In the current study, with a less complex writing task, again we find no effects of observational learning and reflection. This might indicate that the effects of observational learning and reflection in learning to write in the context of academic writing might not be as robust as previously thought.

Conclusion

Constructing sentences is an important skill for writers. When writers are able to fluently combine sentences, this might help them to convey their thoughts and ideas, and positively affect the readability of their texts (Saddler & Graham, 2005). Teaching students to effectively combine sentences might therefore be desirable. The aim of our study was to gain more insight into the effectiveness of instructional methods for teaching sentence-combining, by conceptually replicating Zimmerman and Kitsantas (2002). In addition, by systematically comparing different types of models with and without reflective activities, this study contributes to a growing body of literature on the effects of observational learning within the writing domain.

In our study, learning from a model, either a coping or a mastery model, did not lead to a higher writing revision proficiency than learning from performing actual writing tasks, and neither did reflection. Learning from a model did seem to positively influence self-efficacy initially, but this effect disappeared after learners had a chance to practice their sentence-combining skills. Observational learning also positively affected the way the learners implemented the proposed solution strategy and how useful the learners found the intervention, but these factors were not related to performance. From our findings we conclude that the effectiveness of observational learning and the role of reflection might be susceptible to other factors, which prevents us from deducing clear implications from our findings. While other studies have repeatedly reported on positive effects of observational learning, our current study and the studies presented in Chapters 2 and 3 seem to imply that within our specific context, i.e., university students learning to write, the effectiveness of observational learning has not yet been established.

**

In Chapters 2, 3 and 4, we focused on the effectiveness of observational learning and learning by doing, with and without reflective activities in the cognitive, writing domain. In Chapter 5, we investigate the effects of instructional method and reflection in a completely different domain: motor skill acquisition, which allows us to test to what extent our findings from a cognitive domain transpose to the motor skill domain.

Appendix A – Methodological Similarities and Differences between Zimmerman and Kitsantas (2002) and Current Study

Similarities and adaptions

- The *five-step solution strategy* to solve the writing revision problems was identical in both studies.
- The *writing revision problems* used in the current study were, as in Zimmerman and Kitsantas, selected from Sentence Combining: A Composing Book (Strong, 1981). Since we did not know the exact problems used, we selected problems that could be literally translated into Dutch. Zimmerman and Kitsantas used a total of 25 problems: one example problem, nine problems that the students either studied or observed a model solve, twelve problems the students practiced with in the practice phase, and three problems in the posttest. We used a total of 14 problems: one example problem, four that the students either studied or observed a model solve, three problems in posttest 1 consisted, three problems in the practice phase and three problems in posttest 2. We decided not to include more problems to keep the time spent on the experiment around one hour, which was pretested with five students.
- In Zimmerman and Kitsantas the *modeling* was executed live in the classroom by the experimenter. In the current study we developed a coping model video and an expert model video in order to make sure the students within the same conditions received exactly the same input. Identical to Zimmerman and Kitsantas, the mastery model executed the problems flawlessly, while the coping model started out making several mistakes in a problem, then making only one mistake in a problem and finally executing a problem flawlessly.
- *Writing Revision Skill* was measured twice in both studies, directly following the modeling (and in our case) reflection intervention and following the practice phase. The scoring of the solutions to the problems was identical in both studies. However, the procedure after the modeling intervention differed between the two studies. In Zimmerman and Kitsantas, the students started with the practice problems directly following the modeling intervention. They took the solution of the first practice problem as the measurement of the direct effect of modeling type. In the current study, students completed the first posttest after the modeling and reflection intervention to measure the direct effects of modeling type: they answered the self-efficacy question, completed three writing revision problems, and answered the self-reaction question. This allowed to also measure the direct effects of observational learning versus learning by doing on self-efficacy and self-reaction, which was not the case in Zimmerman and Kitsantas.
- *Self-efficacy* was measured identical to Zimmerman and Kitsantas: the same question, scale and instructions were used. The only difference was that we measured self-efficacy three times (after an example solution, after the modeling and reflection intervention and

after the practice phase), while Zimmerman and Kitsantas measured it twice (after the modeling phase and after the practice phase).
- *Self-reaction* was measured identical to Zimmerman and Kitsantas: the same question, scale and instructions were used. The only difference was that Zimmerman and Kitsantas measured self-reaction once, after the practice phase, while we measured it twice, after the modeling and reflection intervention and after the practice phase.

Additions to the original design
- *Reflection* was added as a factor in order to investigate the role of reflection in observational learning.
- Two dependent variables were added: we added *implementation of the five-step solution strategy* as variable in order to investigate whether modeling type affects the strategies students use when solving problems, which could explain possible differences between conditions. We also included *satisfaction with the modeling intervention* as a variable, to study whether students have a preference for a particular type of learning and whether this is related to performance.

Deletions from the original design
We excluded the independent variable *social feedback* from the design. In this study, we wanted to investigate the role of reflection and in order to not overcomplicate the design, we excluded social feedback. In Zimmerman and Kitsantas' study, social feedback did not interact with modeling type, therefore we did not expect this exclusion to influence the effect of modeling type.

We did not include *intrinsic interest* as a dependent variable in the current study, because we were not able to measure this in a comparable manner, because of differences in the undergraduate curriculum of our participants.

CHAPTER 5

5

Learning How To Throw Darts: Effects of Modeling Type and Reflection on Novices' Dart-Throwing Skills

This chapter is based on:

Van der Loo, J., Krahmer, E., & van Amelsvoort, M. (2021). Learning How to Throw Darts. Effects of Modeling Type and Reflection on Novices' Dart-Throwing Skills. *Journal of Motor Behavior*, *53*(1), 105-116. https://doi.org/10.1080/00222895.2020.17328

Abstract

In this study, we investigated the effects of modeling type and reflection on the acquisition of dart-throwing skills, self-efficacy beliefs and self-reaction scores by conceptually replicating a study by Kitsantas, Zimmerman, and Cleary (2000). Participants observing a novice model were expected to surpass participants observing an expert model who in turn were expected to outperform participants who learned without a model. Reflection was hypothesized to have a positive effect. 156 High school and university students were tested three times: in a pre-test, after a modeling intervention, and after a practice round. Contrary to what was expected, we found no main effects of modeling type and reflection. No interaction effects were found either. There was an effect of testing moment, indicating that participants improved dart-throwing skills, self-efficacy beliefs, and self-reaction scores over time. With these findings, we are not able to replicate Kitsantas et al. From our study, we conclude that observational learning, irrespective of the model's skill level, combined with physical practice, yields similar results as mere physical practice.

Introduction

Acquiring a motor skill, such as throwing darts, is generally not learned from reading a book, but rather by observing others playing darts, by practicing the throws yourself or a combination of observational learning and physical practice. In this chapter on the comparison of observation and physical practice in learning a motor skill, more specifically throwing darts, we address three issues. First, we aim to determine how observational learning, both with and without physical practice, contributes to learning a motor skill compared to physical practice. Second, we address the concerns related to the skills level of the model, by comparing observational learning from a novice model and an expert model with physical practice. Third, we explore the role of reflection by the learner in both observational learning and physical practice. We will do so by conceptually replicating an earlier study by Kitsantas, Zimmerman and Cleary (2000).

Observation versus Physical Practice

Observational learning is the process of learning a new task by watching someone else performing this task. It relies on multiple capabilities: learners should be able to infer the intentions of others from action observation, process others' action outcomes and combine these sources of information in order to select behaviors leading to desired outcomes later on (e.g., Bandura, 1977; Monfardini et al., 2013; Rak, Bellebaum, & Thoma, 2013). With action observation, the learner does not have to generate a simulated representation of the movement, as the key perceptual information is provided in the form of an external stimulus being observed (Ram, Riggs, Skaling, Landers, & McCullagh, 2007). Arguably, this reduces the complexity of the learning task, since the observer is not as cognitively engaged in the task compared to a performer of that same task. Therefore, the observer can more easily break the whole task in subcomponents, selecting essential information and constructing appropriate strategies to reconstruct the task (Cordovani & Cordovani, 2015; Wulf et al., 2010).

Observational learning is generally found to be more effective than doing nothing (e.g., D'Innocenzo et al., 2016) or than verbal instructions alone (Janelle & Hillman, 2003). However, the literature is inconclusive when it comes to comparison of mere practice, mere observation and the combined effects of practice and observation. Some studies imply that observational learning combined with physical practice, leads to more effective ways of performing a task than mere physical practice (Horn, Williams, Scott, & Hodges, 2005; Kitsantas et al., 2000; Shea, Wright, Wulf, & Whitacre, 2000; Weeks & Anderson, 2000), while in other studies it is suggested that there is no advantage of observation combined with physical practice over mere physical practice (Weir & Leavitt, 1990). What complicates the comparisons between these studies, is that in the majority of studies (except for Weir & Leavitt, 1990), physical practice is compared to observation combined with physical practice, confounding the modeling effects with the subjects' practice. In the current study, we therefore aim to determine how observational learning and physical practice contribute to learning a motor skill by systematically comparing observational learning with and without physical practice to mere physical practice.

Type of Model in Observational Learning

Within observational learning the skill level of the model may influence the learning process (Kitsantas et al., 2000; Shea et al., 2000; Wesch, Law, & Hall, 2007). Some have advocated the use of expert models, since they display a standard of reference against which observers are able to detect their own errors and issue appropriate correction, which facilitates constructing a mental representation (Carroll & Bandura, 1987; Ferrari, 1999). Others have advocated the use of novice models, gradually improving their performance, since they provide more information about strategy implementation and error correction than expert models do (Darden, 1997; Ferrari, 1999; Zimmerman & Kitsantas, 2002).

Studies explicitly comparing the effects of novice and expert models in motor skill acquisition have yielded equivocal results. Kitsantas et al. (2000) found that observing a novice model leads to more effective learning of a motor skill than observing an expert model, which in turn leads to better learning than learning without a model, while in a study by Rohbanfard and Proteau (2011) observing an expert model leads to more stable performances with less variability than observing a novice model. Several other studies, on the other hand, found no effects of modeling type on different types of motor skill performance (see e.g., Blanchard, 2014; Blandin, Lhuisset, & Proteau, 1999; Moore, Lelievre, & Ste-Marie, 2019; Pollock & Lee, 1992; Weir & Leavitt, 1990). In these latter studies, physical practice or combinations of physical practice and observation, irrespective of the models' skill level, increased performance.

In view of these contrasting findings, we aim to further explore which model most effectively promotes learning a motor skill by systematically comparing the observation of a novice model, an expert model and observing no model in learning how to throw darts.

Reflection in Observational Learning and Physical Practice

The effects of observational learning are often explained by the fact that it allows learners to construct a mental representation of the desired performance. The observer watches a model and transfers the provided information to his or her own acquisition by judging which parts could be beneficial and how they could be used. It is hereby of importance that observers are capable of identifying relevant features or key movement pattern elements of a motor skill. Adding instructional components to the modeled information could facilitate identifying these relevant cues. However, only a limited number of studies have included such instructional components directing attention to these relevant cues in motor skill learning (Ste-Marie et al., 2012).

In more cognitive research domains, mostly within writing research in which students learn how to write (fairly) complex texts, instructional components are included in the observational learning process. Observers are encouraged to carry out different cognitive, reflective activities, by asking them for example to monitor, evaluate and elaborate on models' performances (e.g., Braaksma, Rijlaarsdam, Van den Bergh & Van Hout Wolters, 2006; Raedts, Van Steendam, De Grez, Hendrickx, & Masui, 2017). These reflective activities support students in developing criteria for effective texts and writing processes which transfer to their own writing, yielding higher quality writing performance, in terms of the degree of selection of relevant and correct information, level and quality of integration of selected ideas, and textual organization (Braaksma, Van den Bergh, & Couzijn, 2001).

From these findings, it can be concluded that reflection arguably supports learners in developing a mental representation of the desired performance and in developing performance strategies which might lead to a more successful performance. In the current study, we therefore explore whether these findings from writing research can be transposed to motor skill learning, by adding reflection to the design.

Aim of the Study

In the current study, we investigate the effect of modeling type (novice, expert and no model) and reflection on acquiring a motor skill. Since previous studies on the effect of modeling type in learning a motor skill vary substantially in their designs and analyses, we chose to conceptually replicate one of the earlier studies, Kitsantas et al. (2000), which allows us to test the robustness of the theoretical implications (Zwaan, Etz, Lucas, & Donnellan, 2018). We chose this particular study because from the handful of studies directly comparing effects of novice and expert models on motor learning, this one resulted in significant differences in the effectiveness of the different types of models, and is conceptually closest to what we aim to explore.

Kitsantas et al. (2000) investigated the effect of modeling type on dart throwing performance and self-regulatory measures with 60 female high school students. The students were first presented with written instructions, followed by either watching a video of a novice model or an expert model, or not watching any video (physical practice condition). The girls were then given fifteen minutes to practice dart throwing, after which they were tested on dart-throwing performance, self-efficacy, error attribution, self-reactions, and intrinsic interest. Kitsantas et al. (2000) found that the girls who had observed the novice model outperformed the girls who had observed the expert model on all measures. The girls who observed an expert model scored in turn better than the girls in the physical practice condition.

In the current study, we closely follow the procedures and measures of Kitsantas et al. (2000) which will be described in the method section. However, we extended the original design in four aspects. First, we add reflection as a factor to the design. In Kitsantas et al. (2000), some reflective activities were present in the novice model condition but absent in the expert model condition and in the physical practice condition, which complicates comparing the different conditions. Second, we extended our sample to not only include high school students but also university students. Third, we add a measure, implementation of dart techniques, to establish whether modeling type influences the physical execution of the performance. Finally, we test the participants three times during the experiment, while participants in the original study were only tested once. This allows us to test possible direct effects of observational learning and combined effects of observation and physical practice.

Because of the replication character of the current study, our hypotheses are similar to Kitsantas et al. (2000). First of all, modeling type is expected to influence the acquisition of a novice motor skill. We hypothesize that (H1a) observing a model leads to a higher performance on all measures than not observing a model, and that (H1b) observing a novice model leads to a higher performance than observing an expert model. Secondly, based on earlier findings concerning the influence of reflection in writing research (Braaksma et al., 2001), we hypothesize that (H2) reflection leads to a higher performance on all measures than no reflection. And thirdly, by adding reflection to physical practice, an interaction effect between modeling type and reflection is hypothesized, in which we expect that (H3) reflection reduces the effect of modeling type on all measures.

Method

Design

A 3 x 2 design was used in this experiment, with modeling type (novice model, expert model and no model) and reflection (yes or no) as the between-subjects factors. This resulted in six conditions to which 156 participants were randomly assigned. The experiment was run with university students and with high school students, in order to increase power and generalizability. In Kitsantsas et al. (2000) the sample size was 60. We therefore aimed at a minimum of 150 participants, based on the suggestion by Simonsohn (2015) that a way to determine sample sizes in replication studies is to take 2.5 times the original sample size.

The participants were tested on dart-throwing skills, self-efficacy, self-reaction and error attribution. We tested dart-throwing skills, self-efficacy and self-reaction three times within the experiment. The first test (pre-test) was at the start of the experiment, so we could establish a baseline. The second test (post-test) was directly after the experimental treatment, in order to measure direct effects of observational learning from a novice model, an expert model, or of physical practice. The third test (delayed post-test) took place after a ten-minute practice session, in order to determine combined effects of observational learning and physical practice. Error attribution was only measured during the delayed post-test. In the experiment with the high school students, we added implementation of the dart techniques as a measure, in order to detect possible differences in the way the darts were thrown. This was measured during the post-test. Table 1 displays which dependent variables were measured in each test.

Table 1.
Dependent Variables Measured in the tests

	Dart Throwing	Self-Efficacy	Self Reaction	Error Attribution	Dart Technique*
Pre-test	+	+	+		
Post-test	+	+	+		+
Delayed post-test	+	+	+	+	

* Dart Technique was only measured for the high school students

Participants

In this study 156 students (88 women, equally distributed over conditions) participated. Ninety of these participants were undergraduate students recruited from a participant pool at Tilburg University (M_age =21,2 years, SD = 3.87). The other 66 participants were Dutch high school students (M_age = 14,3 years, SD = 0.88) who took part in the experiment voluntarily. Table 2 presents an overview of the distribution of the participants over conditions.

Table 2.
Distribution of Participants over Conditions

	Reflection		No Reflection		Total
	University	High school	University	High school	
Novice Model	16	11	15	11	53
Expert Model	15	11	15	11	52
No Model	14	11	15	11	51
Total	78		78		156

Materials

Dart Board. Identical to the study by Kitsantas et al. (2000), we used a dart board with regular concentric circles. The dart board had a bullseye and nine consecutive circles, with each circle having a width of 2.54 centimeter. A numerical value was assigned to each circle, beginning with a value of 10[5] for the bullseye in the center and successively diminishing in assigned values by 1 until the outermost circle, which had a value of 1. The dart board was positioned, in line with the rules of the British Darts Association (2016), at a distance of 2.37 meters of the throwing line, with the bullseye at a height of 1.73 meters from the ground.

5 In Kitsantas et al. (2000) the dart board consisted of seven concentric circles, which resulted in a maximum average dart-throwing score of 7.

The participants were provided with six regular darts so they could make six consecutive throws during the tests.

Hand-out Dart Instructions. The participants were provided with a hand-out which contained instructions for dart throwing. The instructions included information on five subskills: grip, stance, sighting, throw and follow through, and were derived from the instructions used in Kitsantas et al. (2000) and translated into Dutch. Table 3 provides an overview of the instructions.

Table 3.
Dart-Throwing Instructions provided to the participants (taken from Kitsantas et al.,2000)

Subskill	Instruction
Grip	Hold the dart between your first and second finger and the thumb. Simply grasp the dart comfortably.
Stance	Stand behind the white throwing line facing the target. Stand comfortably with your feet slightly apart. If you are right handed, the right foot should be slightly ahead of the left, touching the toe line and pointing toward the board. If you are left handed, place the left foot forward.
Sighting	Keep your arm close to your body. Using your arm and wrist, and with the elbow acting as a fulcrum, bring the dart back toward your face it until almost brushes your cheek where you find it most comfortable to stop.
Throw	Keep all the other parts of your body still when you throw. Your head must be held steady, and you must not jerk the throw. Try to develop a smooth, vertical throw using the wrist and elbow as pivots. Hold your elbow steady, and keep it parallel to the floor. Your wrist should be loose and laid slightly back. Use only the forearm and wrist to throw in a vertical motion. The throw need not be hard, but it must be crisp. The dart should get to the board quickly with as little trajectory as possible.
Follow-through	After you release the dart, simply allow your arm to continue in its natural motion. Let your hand, with your fingers fully extended, follow the dart as it moves toward the target.

Note. The instructions were translated into Dutch before providing them to the participants.

Videos. Two different videos were recorded, one of a novice model playing darts (see Figure 1) and one of a expert model playing darts. The same male model appeared in both videos. The videos displayed information on all five subskills described on the hand-out and in each of them the model threw 15 darts. In the novice model condition, the model started out by making several errors in his dart-throwing technique. These errors were made by not following the instructions to dart-throwing that were described, for instance by holding the dart with all five fingers instead of the three mentioned in the instruction. The model commented on his own performance and gradually improved his dart-throwing skills. In the expert model video, the model started out with a flawless technique, and maintained this throughout the video. The model commented on his own performance, by describing the techniques he was using. Both videos lasted approximately 1 minute and 40 seconds.

Figure 1.
Still of the Model Playing Darts

Reflection Questions. Participants in the reflection conditions were asked to answer reflection questions on each of the five subskills from the instructions. Given the differences in content between the conditions, reflection questions also differed somewhat per condition. In the expert-model with reflection condition, participants were asked open-ended questions for each subskill. An example for the subskill stance is: "How did the dart-thrower position his feet?". In the novice-model with reflection condition, participants were asked to answer additional questions for each subskill, namely "What kind of mistakes did the model make?" and "How did the model correct his mistakes?". In the no-model with reflection condition, participants reflected on their personal performance, for example "How did you position your feet?". They were then asked what kind of mistakes they made (if any), and how they would correct these the next time. These questions allowed participants to reflect on and process effective and/or less effective strategies in all conditions.

Measures

Dart-throwing skills. To measure dart-throwing skills, we added up the points for all six consecutive throws. The minimum score per dart was zero and the maximum score was ten. The average dart-throwing score per dart was calculated by dividing the total score by six. Dart-throwing skill was measured in all three tests.

Self-efficacy. Before each dart-throwing test, participants were asked how confident they felt that they could (1) throw 9 points with one dart, (2) throw 7 points with one dart, (3) throw 5 points with one dart, and (4) throw 3 points with one dart. Their scores were measured on a scale from 0 to 100 with 10-point intervals, with 0 being not sure and 100 being very sure. This self-efficacy measure was comparable to the measure used by Kitsantas et al. (2000). For each of the three tests, the average score of these four questions was used as the participants' self-efficacy score (Cronbach's alpha in all three tests > .86).

Self-reaction. Participants had to indicate for each of the three tests on a scale from 0 to 100 how satisfied they were with their own performance, with 0 being not satisfied and 100 being very satisfied. This measurement was identical to that of Kitsantas et al. (2000).

Error Attributions. After the delayed post-test, we asked the participants why they thought they had missed the bullseye. If they hit the bullseye, they could skip this question: this was the case for five participants. Their answers were grouped into six categories, namely: type of strategy, amount of effort, level of ability, amount of practice, I don't know, and other. This measurement was only carried out after the last test, in line with Kitsantas et al. (2000).

Dart Techniques. In the experiment with the high-school students, we added the measure dart techniques. In order to measure whether the participants implemented the correct dart technique, we filmed the dart-throwing test of the high-school students' post-test, which followed the experimental treatment immediately. We scored the first and last throw on the five subskills presented in the instructions: grip, stance, sighting, throw and follow through. We scored 1 point if the subskill was implemented correctly, and 0 if it was not implemented correctly. This resulted in a possible minimum score of 0 and a maximum score of 10 (5 for the first throw and 5 for the last throw).

Procedure

The university students were individually taken into a room where they were introduced to the procedure by an experimenter and filled out a consent form. The high school students were taken into a large physical education room in pairs. In the room, two experimenters were available, one for each student. The high school students were each taken to opposite sides of the room and introduced to the procedure individually by one of the two experimenters. Parental consent was obtained by the managing director of the high school, prior to the experiment. During the experiment the high school students did not face each other, since we set up the room in such a way that the two dart boards were on opposite walls of the room. This way the students had their backs turned to each other, with at least 3-5 meters between them. All participants then filled out a form with demographic information, and they performed a baseline test in which their self-efficacy, dart-throwing skills, and self-reaction were measured (pre-test). Within this test, participants first answered the question about self-efficacy, then they threw six consecutive darts, and after throwing the darts, they answered the self-reaction question.

After the pre-test, they were provided with the hand-out containing the instructions to dart throwing (grip, stance, sighting, throw, and follow through). The participants were asked to study these until they felt confident that they knew what to do. After having read the instruction, the procedure depended on the condition participants were assigned to.

No model, no reflection. Participants were tested again straight after reading the instructions (post-test). Then, they got to practice dart throwing for ten minutes. After the ten-minute practice, the participants were tested for one last time (delayed post-test).

Novice model, no reflection. Participants were shown a video of the model who started out by making certain mistakes, but significantly improved during the video. After having watched the video, the participants performed the same test again, thus, their self-efficacy, dart-throwing skills, and self-reaction were measured (post-test). Then, they got to practice dart throwing for ten minutes. After the ten-minute practice, the participants were tested for one last time (delayed post-test).

Expert model, no reflection. Participants were shown a video of a dart-thrower who made 15 good throws. After having watched the video, the participants performed the same test again, thus, their self-efficacy, dart-throwing skills, and self-reaction were measured. Then, they got to practice dart throwing for ten minutes. After the ten-minute practice, the participants were tested for one last time (delayed post-test).

No model, with reflection. Participants were asked to perform their first fifteen practice throws. After that, they were asked to reflect upon their own performance according to the five subskills described in the instructions (grip, stance, sighting, throw, and follow through). After the reflection, they were tested again on self-efficacy, dart-throwing skills, and self-reaction (post-test). Afterwards, they continued to practice. Because we wanted to make sure that the total practice time was equal across all conditions, the duration of the first fifteen throws of the participants in the no model, with reflection group, was part of their total ten minutes of practice. Thus, if participants performed their first fifteen throws in three minutes, they had an additional seven minutes to practice after the reflection. After the practice period, the participants were tested for one last time (delayed post-test).

Novice model, with reflection. After having read the instructions, participants were shown the video of a dart-thrower who started out by making certain mistakes, but significantly improved over the rounds, the model threw 15 darts in total. Participants were asked to evaluate the model’s performance on the five subskills explained in the instruction and to report mistakes and possible improvements. After having watched the video and having reflected on the model’s performance, the participants performed the same test again, thus, their self-efficacy, dart-throwing skills, and self-reaction were measured. Then, they got to practice dart throwing for ten minutes. After the ten-minute practice, the participants were tested for one last time (delayed post-test).

Expert model, with reflection. After having read the instructions, participants were shown the video of a dart-thrower who made 15 good throws. They were asked to evaluate the performance of the model on the subskills explained in the instruction. After having watched the video and having reflected on the model's performance, the participants performed the same test again, thus, their self-efficacy, dart-throwing skills, and self-reaction were measured. Then, they got to practice dart throwing for ten minutes. After the ten-minute practice, the participants were tested for one last time (delayed post-test).

Analyses

A repeated measures multivariate ANOVA was conducted, for each dependent variable (dart throwing skills, self-efficacy and self-reaction), with modeling type (no model, expert model, novice model), reflection (absent, present) and group (university students, high school students) as between subjects-variables and test moment (pre-, post- and delayed post-test) as a within-subjects variable. Mauchly's test for sphericity was used to test for homogeneity of variance. We applied Bonferroni corrections for the pairwise comparisons. For attribution, we performed a chi square analysis, since attribution was measured on a categorical score. An ANOVA with modeling type, reflection and groups as independent variables was conducted for measuring implementation of dart techniques.

Results

Dart-throwing skills

A within-subjects main effect of testing moment was found for dart-throwing skills ($F(2,144) = 56.01$, $p < .01$, $\eta p2 = .28$), with the mean scores gradually increasing each testing moment ($M_{pre} = 4.40$, CI = [4.13, 4.68], $M_{post} = 4.47$, CI = [4.21, 4.74], $M_{delay} = 5.62$, CI = [5.40, 5.85]). The mean difference between the pre-test and the delayed post-test (M-diff = -1.22) was significant, $p < .01$, CI = [-1.48, -.96] and so was the difference between the post-test and the delayed post-test (M-diff = -1.15), $p < .01$, CI [-1.40, -.90]). However, the difference between the pre-test and post-test (M-diff = -.07) turned out not to be significant, $p = .59$, CI = [-.33, .19].

Group had a significant main effect on dart-throwing skill, $F(1,144) = 16.59$, $p < .01$, $\eta p2 = .10$, with the university students (M = 5.27) scoring significantly higher than the high school students (M = 4.40).

No significant main effect of modeling type on dart-throwing skills was found, $F(2,144) = 2.13$, $p = .10$. The data numerically suggest that best results are obtained by observing an expert model (M = 5.10, CI = [4.73 , 5.47]), followed by a novice model (M = 4.86, CI = [4.51, 5.22]) and no model (M = 4.54, CI = [4.17, 4.90]).

Reflection had no main effect on dart-throwing skill either, $F(1,144) = 0.08$, $p = .77$. The dart-throwing scores in the conditions without reflection ($M = 4.86$, CI = [4.57, 5.16]) were not significantly different from those in the conditions with reflection ($M = 4.80$, CI = [4.51, 5.10]).

In addition, none of the interaction effects were found to be significant ($Fs < 1.43$, $ps > .24$). Table 4 displays the mean dart-throwing skills scores per condition.

Table 4.
Average Dart-throwing Skills (with SDs) as a function of Reflection (Reflection, No reflection), Model (No Model, Expert, Novice) and Test (pre-, post-, delayed post-test)

	No Model			Expert			Novice		
Test	Pre	Post	Delayed	Pre	Post	Delayed	Pre	Post	Delayed
No reflection	4.0 (1.8)	4.2 (2.0)	5.8 (1.5)	4.7 (1.8)	4.8 (1.6)	5.9 (1.6)	5.0 (1.6)	4.7 (1.7)	5.6 (1.5)
Reflection	3.9 (1.4)	4.3 (1.5)	5.4 (1.6)	4.8 (1.4)	5.2 (1.7)	5.9 (1.3)	4.4 (2.2)	4.3 (1.9)	5.6 (1.7)

Self-efficacy

The pattern for self-efficacy is very comparable to that of dart-throwing skills. Again, a main effect of testing moment on self-efficacy was found, $F(1.77, 254.48) = 185.40$, $p < .01$, $\eta p2 = .56$. Self-efficacy scores on the pre-test ($M = 49.93$, CI = [47.22, 52.63]), the post-test ($M = 59.78$, CI [57.17, 62.39]), and the delayed post-test ($M = 68.56$, CI = [66.01, 71.11]) were significantly different. All the individual differences were significantly different at a p-value of < .01.

Group also had a significant main effect on self-efficacy, $F(1,144) = 49.68$, $p < .01$, $np2 = .26$, with the university students ($M = 50.94$) scoring significantly lower than the high school students ($M = 67.90$).

The main effect of modeling type did not yield significance, $F(2,144) = 2.73$, $p = .07$. Again, the data numerically suggest that best results are obtained by observing a expert model ($M = 63.42$, CI [59.25, 67.59]), followed by observing a novice model ($M = 57.64$, CI [53.60, 61.68]) and no model ($M = 57.21$, CI [53.06, 61.35]). However, this effect failed to reach significance.

The main effect of reflection was not found to be significant either, $F(1,144) = 3.62$, $p = .06$. Similar to the effect of reflection on dart-throwing skills, the data numerically suggest that highest scores are obtained by not reflecting ($M = 61.71$, CI = [58.35, 65.07]) instead of reflecting ($M = 57.13$, CI = [53.77, 60.50]). Again, no significant interaction effects were found ($Fs < 0.59$, $ps > .55$). In Table 5 the mean self-efficacy scores per condition are displayed.

Table 5.
Average Self-Efficacy (with SDs) as a function of Reflection (Reflection, No reflection), Model (No Model, Expert, Novice) and Test (pre-, post-, delayed post-test)

	No Model			Expert			Novice		
Test	Pre	Post	Delayed	Pre	Post	Delayed	Pre	Post	Delayed
No reflection	50.6 (20.0)	56.7 (18.2)	67.3 (18.2)	51.5 (16.9)	66.0 (15.7)	75.1 (13.7)	49.5 (19.6)	60.0 (17.8)	66.7 (16.2)
Reflection	43.5 (21.3)	52.8 (19.3)	64.0 (18.5)	49.6 (17.7)	60.4 (17.4)	68.9 (65.1)	44.7 (23.4)	55.3 (19.6)	63.5 (18.0)

Self-reaction

The pattern of results for self-reaction were very similar to those of self-efficacy and dart-throwing skills. A significant main effect of testing moment on self-reaction scores was found, $F(2, 144) = 14.93$, $p < .01$, $\eta p2 = .09$. In the post-test the scores dropped compared to the pre-test ($M_{pre} = 63.23$, C.I. [60.31, 66.15], $M_{post} = 56.39$, CI = [52.94, 59.84]. In the delayed post-test ($M_{delay} = 67.14$, CI = [64.07, 70.21]) the mean scores increased again. The mean difference between the pre-test and 2 (M-diff = 6.84) was significant, $p < .01$, CI = [1.72, 11.96] and so was the mean difference between the post-test and delayed post-test (M-diff = -10.74), $p < .01$, CI [-15.68, -5.81]). However, the mean difference between the pre-test and delayed post-test (M-diff = -3.91) turned out not to be significant, $p = .10$, CI = [-.8.28, .47].

No significant main effect of modeling type, ($F(2,144) = 2.44$, $p = .09$) or reflection ($F(1,144) = 1.54$, $p = .22$) was found, indicating that no model (M = 59.85, CI [56.04, 63.67]), observing a expert model (M = 65.65, CI [61.81, 69.50]) and observing a novice model (M = 61.25, CI [57.53, 64.97]) do not lead to significantly different self-reaction scores, and neither do not reflecting (M = 63.63, CI [60.53, 66.72]) and reflecting (M = 60.88, CI [57.78, 63.97]). Again, the numerical differences suggest best results are obtained by observing an expert model, followed by a novice model and learning by doing, and by not reflecting instead of reflecting.

Group had no effect on self-reaction, $F(1,144) = 2.61$, $p = .11$ In addition, no significant interaction effects were found, $Fs < 1.54$, $ps > .22$. See Table 6 for an overview of the self-reaction scores.

Table 6.
Average Self-reaction Scores (with SDs) as a function of Reflection (Reflection, No reflection), Model (No Model, Expert, Novice) and Test (pre-, post-, delayed post-test)

	No Model			Expert			Novice		
Test	Pre	Post	Delayed	Pre	Post	Delayed	Pre	Post	Delayed
No reflection	63.9 (17.0)	56.9 (25.1)	71.1 (19.9)	66.5 (19.8)	58.6 (14.1)	73.2 (15.6)	65.5 (17.7)	56.2 (19.7)	63.3 (18.9)
Reflection	54.8 (21.9)	52.5 (22.3)	60.5 (15.1)	66.8 (13.9)	62.5 (23.1)	67.8 (21.9)	64.9 (18.5)	54.4 (22.4)	66.6 (22.0)

Attribution

Table 7 shows the frequency of each answer for the different modeling conditions, whereas these frequencies are displayed for the reflection conditions in Table 8.

Table 7.
Attribution (Count and Percentage) as a Function of Modeling Type

	Attributional source					
Group	Strategy	Effort	Ability	Practice	Don't know	Other
Control	6 (12%)	3 (6%)	20 (39%)	17 (33%)	3 (6%)	2 (4%)
Expert	7 (15%)	3 (6%)	19 (39%)	14 (29%)	1 (2%)	4 (8%)
Novice	9 (17%)	3 (6%)	25 (48%)	11 (21%)	1 (2%)	3 (6%)

Table 8
Attribution (Count and Percentage) as a Function of Reflection Group

	Attributional source					
Group	Strategy	Effort	Ability	Practice	Don't know	Other
No Reflection	10 (14%)	2 (3%)	29 (39%)	25 (34%)	5 (7%)	3 (4%)
Reflection	12 (15%)	7 (9%)	35 (46%)	17 (22%)	0 (0%)	6 (8%)

Before the chi-square analyses could be performed, we recoded the answers into four categories, combining Effort, Other and Don't Know in order for all cells to have a count above five, which is required for a chi-square analysis. The analysis showed no significant differences between the modeling groups, α (6) = 2.75, $p = .84$ and between the reflection groups, α (3) = 2.60, $p = .46$. These results indicate that there is no significant difference between participants who learned from observing an expert model, a novice model or no model, and between those who reflected and those who did not in what they attribute their errors to.

Implementation of Dart Techniques

In the experiment with the high-school students, we measured whether the participants implemented the dart techniques in a correct manner. The ANOVA revealed no significant effects of modeling type, $F(2,53) = 0.64$, $p = .53$ or reflection, $F(1,53) = 0.07$, $p = .80$. No interaction was found either, $F(2,53) = 0.57$, $p = .57$. These results indicate that modeling type and reflection did not influence how participants used the techniques when throwing darts. An overview of the mean scores is displayed in Table 9.

Table 9
Scores on Dart Techniques Implementation (SDs) as a function of modeling type and reflection (min. score 0, max. score 10).

	No Model	Expert Model	Novice Model	Total
Reflection	6.9 (1.8)	7.3 (1.7)	6.8 (1.9)	7.0 (1.9)
No Reflection	6.5 (1.8)	7.3 (2.1)	7.6 (1.6)	7.1 (1.7)
Total	6.7 (1.8)	7.3 (1.8)	7.2 (1.8)	

Correlations

To further explore the relationships between dart-throwing skills, self-efficacy and self-reaction, we used a Pearson's correlation test. In the pre-test, the results show a weak but significant correlation between self-efficacy and dart-throwing skills, $r = .29$, $n = 156$, $p < .01$., and a medium strong correlation between self-reaction and dart-throwing skills, $r = .47$, $n = 156$, $p < .01$. The correlation between self-efficacy and self-reaction, however, was very weak and non-significant, $r = .05$, $n = 156$, $p = .52$.

For the post-test, the correlations are very similar to the pre-test. Again, the correlation between self-efficacy and dart-throwing skills is weak but significant, $r = .24$, $n = 156$, $p < .01$, and the correlation between self-reaction and dart-throwing skills is of medium strength and significant, $r = .51$, $p < .01$. The correlation between self-efficacy and self-reaction is again not significant and weak in this second test, $r = .05$, $n = 156$, $p = .54$.

In the delayed post-test, again we find a weak but significant correlation between self-efficacy and dart-throwing skills, $r = .28$, $n = 156$, $p < .01$, and a medium strong correlation between self-reaction and dart-throwing skills, $r = .57$, $p < .01$. However, in the delayed post-test, we do find a weak but significant correlation between self-efficacy and self-reaction, $r = .22$, $n = 156$, $p < .01$.

General Discussion

Observational learning is often included in learning a complex motor skill, such as dart throwing. By action observation of a model performing the motor skill, learners are arguably less cognitively involved in the task, which could support them in constructing a mental representation of the action, leading to a higher or more accurate performance. Since the findings in previous research are inconclusive on the effects of different modeling types, in this study we sought to determine what would be the most effective way of learning a motor skill by systematically comparing observational learning from a novice model, observational learning from an expert model and learning from physical practice. In addition, we explored whether adding reflective activities to both observational learning and physical practice could enhance learning.

Modeling type: expert model and novice model versus physical practice

We found that observational learning from a novice model, an expert model and learning from physical practice yielded similar results on dart throwing skills, self-efficacy beliefs, self-reaction, error attribution, and implementation of dart techniques, which means we have not found support for H1a an H1b. This was the case in both the post-test, directly following the modeling intervention, and the delayed post-test, following the practice session[6].

In the post-test, participants who watched one of the models, had similar darts scores as those who practiced, and they also reported similar levels of self-efficacy. It should be noted, however, that dart throwing performance did not change significantly from the pre-test to the post-test, so it could also be argued that a short practice period or mere observational learning does not suffice for learning to throw darts. Another interesting find is that in all groups satisfaction dropped significantly from the pre-test to the post-test, while the dart scores did not change. The participants were thus less satisfied with the same score in the post-test, which implies that the learners thought they had learned something from either watching the model or the short practice period. In the post-test, we also measured how accurate the high school students implemented the dart techniques presented to them. Again, no differences were found between conditions. Observational learning did not lead to a more accurate dart throwing technique. A limitation is that we only measured implementation of dart techniques in this post-test. This did allow us to determine immediate effects of observational learning compared to physical practice. However, this prevents us from making statements on the combined effects of observational learning on the implementation of dart techniques.

The results of the delayed post-test, again reveal no differences between observational learning combined with physical practice and mere physical practice. After a ten-minute practice session, all groups significantly improved their dart throwing performance. In addition, all groups showed more confidence in themselves and were more satisfied with their performance, compared to the pre-test and post-test, irrespective of condition. This was also the case for error attribution.

In the current study we were not able to replicate the findings by Kitsantas et al. (2000) who found clear effects of modeling type, even though our experimental procedures and measurements were very similar. Our test was also sufficiently powered: our sample was larger than in Kitsantas et al. (respectively 156 and 60) and the test was sensitive enough to find statistical differences for testing moment, indicating that the participants did learn from the interventions. Furthermore, we also found no indications, contrary to Horn et al. (2005), that observational learning facilitates adopting motion information, leading to a more accurate performance.

6 To quantify the evidence of the absence of the effect, we performed post-hoc equivalence tests (Lakens, 2017) comparing the modeling conditions for all dependent variables. The equivalence tests show that the largest equivalence bounds (based on Cohen's D) for dart performance, self-efficacy, self-reaction and implementation of dart techniques vary between .48 and .78, meaning that with current sample size, we can similarly reject medium to large and large effects.

Our results are, however, in line with Blanchard (2014), Blandin et al. (1999), Pollock and Lee (1992), and Weir and Leavitt (1990) who also found no differences between modeling types in learning a motor skill. The volatile results on the effects of modeling type on motor skill acquisition frustrate drawing clear conclusions on the implications of our study. Our study and these latter studies seem to indicate that observational learning, of either a novice or expert model, does not positively affect learning compared to physical practice, but it also seems not to hinder learning a motor skill. This seems to be especially the case when it is combined with a physical practice period in which observers can experience sensory feedback which arguably improves muscle control (Blandin et al, 1994; Cordovani et al., 2015) leading to a higher performance.

A limitation to our study is arguably the way we measured dart performance. Since our study was a replication of Kitsantas et al. (2000), we measured dart performance in the same way as they did, to facilitate comparison of the results. However, other researchers, including Hancock, Butler, and Fischman (1995) and Fischman (2015) have argued that with this type of assessment of dart performance is not optimal, because it does not allow determining performance variability, which is considered to be an important characteristic of motor skill learning. Using two-dimensional error measures could support quantifying accuracy, bias and consistency of performance. Future studies should consider adding these two-dimensional error measures to the design. Another limitation is that we did not include a mixed observation group in the design. There are some indications that observing both a novel and expert model might lead to a more stable performance (Rohbanfard & Proteau, 2011), which could be explored in future studies.

Reflection

Previous studies have suggested that integrating instructional components to a motor skill learning regiment could enhance learning, since these components focus attention to relevant movements (e.g., Ste-Marie et al., 2012). In our study, we systematically compared reflecting to not reflecting in both observational learning with a novice and expert model and in physical practice.

In the current study, reflection did not affect dart throwing skills, self-efficacy beliefs, self-reaction, error attribution, and implementation of dart techniques, which means H2 and H3 are not supported by our results. Participants who reflected did not outperform those who did not[7]. It is worth noting that reflection in the different learning conditions was of necessity subtly different: in observational learning the students reflected on the model's performance

7 To quantify the evidence of the absence of the effect, we performed post-hoc equivalence tests (Lakens, 2017) comparing the reflection conditions for all dependent variables. The equivalence tests show that the largest equivalence bounds (based on Cohen's D) for dart performance, self-efficacy, self-reaction and implementation of dart techniques vary between .34 and .53, indicating that with our sample size, we can reject medium and large effects.

while in the physical practice condition, they reflected on their own performance, and this difference might conceivably influence the effectiveness of reflection.

Since we found no effects of reflection, it could be argued that the positive effects of reflection found in observational learning within cognitive domains, such as writing, do not transpose to the motor skill domain. The reflection questions did direct attention to relevant features of the performance, but only after the participants' own performance or that of the models. This means that they had to reflect on all five subskills in hindsight, while compared to the writing domain, the subskills are executed within a very limited time span. This might have affected the effectiveness of the reflection questions.

Conclusion

In the current study, we investigated how observational learning from a novice model or an expert model influence dart throwing skills, self-efficacy beliefs, self-reaction, error attribution and implementation of dart techniques, compared to learning from physical practice. In addition, we sought to explore how reflection affects acquiring a motor skill, and what the interplay is between reflection and observational learning from a novice model, an expert model or physical practice. In our study, it did not matter how someone was taught how to play darts, as long as they practiced, their scores improved, they felt more confident, and they were more satisfied. Our study contributes to the growing body of research on observational learning within the motor skill domain. From these studies, including the current one, no clear view can be established on the effects of observational learning of either a novice or expert model, whether it is combined with physical practice or not, compared to mere physical practice.

CHAPTER 6

Discussion: Learning to Write an Academic Text is Complex!

Discussion: Learning to Write an Academic Text is Complex!

In the research presented in this dissertation, we set out to investigate effective instructional methods for academic writing. One genre that is common in academic writing is a research synthesis: students, for example, are asked to write literature reviews or introductions to research papers in which they have to organize and synthesize information from multiple sources (Boscolo, Arfé, & Quarisa, 2007). Students struggle with this type of task: for example, they have a tendency to organize the sources chronologically instead of by content. This often results in a text in which each paragraph represents one source and in which connections between the different sources are lacking (Granello, 2001). This is illustrated in this first example of a student paper on the effect the number of followers a social media influencer has on the influencer's credibility[8].

The study by Jin and Phua (2014) examined the effect of having many followers on Twitter on the credibility of celebrities. This research shows that when the number of followers is higher, people think they are more reliable and seen as more skilled and handsome.

De Veirman, Cauberghe and Hudders (2017) investigated the effect of the number of followers an influencer has on the attitude towards the influencer. They found that more followers with an influencer led to a more positive attitude towards the influencer.

In this example, we see that the student discusses two of the sources in two separate paragraphs, without establishing any connections between the sources. However, lecturers expect students to identify and verbalize the relations, as can be seen in the next example. This student uses an introductory sentence to make a more explicit link with the topic of the paper, and also uses sign-posting language, such as 'for example', and 'however', to establish the connection between the sources.

One factor that affects the credibility of influencers is the number of followers. For example, Jin and Phua (2014) found that consumers perceived celebrities on Twitter with a lot of followers as more credible than those with only few followers. However, according to De Veirman, Cauberghe and Hudders (2017), a large number of followers, does not directly result in more credibility. Even though influencers were seen as more popular if they had a lot of followers, they were not necessarily seen as more credible.

How can we teach students to write such a well-organized complex text? We know that this is not easy: when learning to write students are often confronted with a double task, writing and *learning* to write, which is arguably very demanding from a cognitive perspective (Braaksma, Rijlaarsdam, Van den Bergh & Van Hout-Wolters, 2008). Previous studies have therefore

8 Translated from Dutch

investigated the effectiveness of observational learning, compared to learning by doing, with different types of audiences and different types of products in learning to write. The premise in this type of research is that observing models lowers the cognitive load, compared to engaging in writing exercises, which would arguably allow learners to construct a mental representation of the performance without having to engage in writing activities themselves, leaving more cognitive attention for actual learning. Multiple studies have reported positive effects of observational learning on performance. However, the direct effects of observational learning within the context of writing complex texts at university level had been relatively understudied. The main aim of this dissertation was therefore to compare the effectiveness of observational learning compared to learning by doing with university students mastering the art of academic writing, which led to our first research question:

Research Question 1: What is the effect of observational learning on text quality and self-efficacy in the context of academic writing compared to learning by doing?

In addition, we explored the role of reflection in learning to write, since in many studies comparing observational learning to learning by doing, reflective activities are present in observational learning while being absent in learning by doing, which possibly confounds the effects of observational learning. This led to the second research question:

Research Question 2: How does reflection affect observational learning and learning by doing in academic writing?

In this dissertation, we aimed at contributing to a growing body of evidence-informed instructional writing practices. Therefore, we attempted to replicate previous studies reporting effects of observational learning. We studied the same variables and effects as in previous studies, following their methods and procedures closely, even though the exact operationalizations differed. Since each chapter already contained a thorough discussion of the individual findings in that study, we shall be relatively brief here. In this chapter, we will first answer the two main research questions. Then the theoretical and practical implications of these findings will be discussed, followed by the overall conclusion.

Research Question 1: Observational Learning versus Learning by Doing

From the four studies conducted in this dissertation we conclude that there are no differences between the effectiveness of observational learning and learning by doing on performance and self-efficacy beliefs. We find this consistently with different types of tasks and in different contexts.

In Chapters 2 and 3 the absence of the effect was found for a complex writing task, i.e. the introduction to a research paper in which students had to synthesize information from multiple sources. The two studies were conducted in an actual academic writing course at Tilburg University. Students who learned via models did not outperform those who learned by performing actual writing assignments. In Chapter 2 we did find some indications that writing preference influenced the effectiveness of observational learning: students with a revising preference seem to benefit more from observational learning than students with a planning preference, but we were not able to replicate this finding in Chapter 3. We also found that students who learned by doing were more satisfied with the instructional intervention: they found the interventions more useful and helpful than students who learned from observation.

With a less complex writing task, i.e., sentence-combining, in the more controlled experimental setting of Chapter 4, observational learning and learning by doing also yielded similar results. The quality of the sentences the students wrote was comparable, and so were the self-efficacy beliefs. We did find that observational learning led to a more accurate execution of specific performance processes, but this did not affect the quality of the performance. Contrary to what we found in Chapter 3, students seemed to prefer observational learning over learning by doing, even though this did not affect performance or self-efficacy.

Since we did not find differences between the effectiveness of observational learning and learning by doing in cognitive, writing tasks, we looked at a motor skill task, throwing darts, in Chapter 5. Again, the type of instruction did not influence performance scores. Even though all students improved their dart scores over time, this was not affected by the type of instruction they received, implying that practice, and not observation, led to the improved performance. We did find that observational learning led to an initial increase of self-efficacy beliefs, but this effect disappeared after a practice period.

This leads us to conclude that within the context of our studies, observing models does not contribute more to learning to write than learning by doing: both instructional methods seem to lead to an equal performance.

Research Question 2: The Role of Reflection

We investigated the role of reflection in learning to write in three studies: with the research synthesis task in Chapter 3, the sentence-combining task in Chapter 4, and the dart-throwing task in Chapter 5. In all three studies, we systematically compared observational learning with and without reflection to learning by doing with and without reflection. In observational learning, reflection was prompted by asking students reflective questions on the behavior of the models they observed, while in learning by doing the students were asked similar questions on their personal performance. In the three studies, we consistently find no differences between the presence and absence of reflection. Adding reflective activities to both observational learning and learning by doing has no effect on writing and motor skill performance nor on self-efficacy beliefs. In Chapter 3 we did find that students preferred reflecting over not reflecting, but this was not related to performance or self-efficacy. We therefore conclude that in the context of this dissertation, reflection has no impact on learning to write in an academic context.

Combined Analysis: Are there overall effects of instructional method and reflection?

In the research from this dissertation, we were not able to replicate previous findings on observational learning or reflection. Why is this? Potentially, several factors could explain this lack of effects. It is conceivable, first of all, that our studies were not sufficiently powered to detect subtle effects of method or reflection, even though our sample sizes in the individual studies were comparable or even larger than those in the original studies we set out to replicate. Given that our studies have a comparable design, i.e., always comparing observational learning with learning by doing, and in studies 2, 3 and 4 comparing the presence or absence of reflection, we can actually address this by performing a combined analysis over multiple studies. In this way, we increase statistical power, thereby enabling us to possible detect more subtle effects of instructional method or reflection.

For this combined analyses, first standardized z-scores were computed for the performance and self-efficacy scores of the final posttest of each experiment, to make sure that differences in scoring between experiments could be statistically compared. Because reflection and self-efficacy were not included in Experiment 1 (Chapter 2), we performed three separate ANOVAs (see Table 1 for an overview). In all ANOVAS we included Experiment as a factor, and we checked for possible interactions between instructional method, reflection and experiment. Below we focus on the findings regarding instructional method and reflection, since experiment never significantly affected performance, and none of the interactions were significant.

6

Table 1.
Overview of the ANOVAS performed in the Combined Analyses

ANOVA	Independent Variable(s)	Dependent Variable(s)	Included Experiments	Total N
1	Instructional Method, Experiment	Final Posttest Performance	1, 2, 3, 4	518
2	Reflection, Experiment	Final Posttest Performance	2, 3, 4	373
3	Instructional Method, Reflection, Experiment	Final Posttest Self-Efficacy	2, 3, 4	373

First, we analyzed the effect of Instructional Method (observational learning versus learning by doing) on performance. There was no significant effect of Instructional Method on Performance, $F(1,517) = 0.59$, $p = .44$. This combined analysis thus does not seem to yield an effect of instructional method on performance either. To quantify the evidence of the absence of the effect of Instructional Method, we performed post-hoc equivalence tests

(Lakens, 2017). Based on the sample size, we set the equivalence bounds at a benchmark of a positive or negative difference of Δ= 0.3, as the smallest effect size of interest. The TOST procedure for Welch's t test for independent samples, revealed that the effect of instructional method observed in the combined analysis was statistically equivalent, because the larger of the two *p* values was less than .05, $t(502) = 2.41$, $p = .008$.

In the second ANOVA, the effect of Reflection (absent or present) on performance was analyzed (experiment 2, 3, 4). The ANOVA revealed that Reflection did not affect performance, $F(1,372) = 2.19$, $p = .14$. The TOST procedure for Welch's t test for independent samples, with equivalence bounds of $\Delta_L = -0.3$ and $\Delta_U = 0.3$, showed that the effect of Reflection was statistically equivalent, because the larger of the two *p* values was less than .05, $t(370) = 2.18$, $p = .015$.

In the third ANOVA, we analyzed the effect of Instructional Method and Reflection on self-efficacy. No main effects of Instructional Method, $F(1,369) = 0.60$, $p = .44$, nor Reflection, $F(1,369) = 0.01$, $p = .94$ were found on self-efficacy. The TOST procedure for Welch's t test for independent samples, with equivalence bounds of $\Delta_L = -0.3$ and $\Delta_U = 0.3$, showed that the effects observed were statistically equivalent, because the larger of the two *p* values for Instructional Method was less than .05, $t(324) = 2.00$, $p = .023$, and so was the larger *p* value for Reflection, $t(367) = -2.53$, $p = .006$.

In short, the results of these combined analyses confirm the findings from the individual chapters: even when we combine the data from experiments, to increase the statistical power of our studies, we find no evidence for effects of instructional method or reflection. In other words, if these effects exist, they must be very small at best.

Explaining the Null Results

With the results of the combined analysis, it seems rather unlikely that our null results can be explained by a lack of power. Another possible explanation for the lack of effects could be the context of our studies: university students learning to write, while much of the evidence of benefits for learning by observation comes from younger learners. Even though Raedts, Rijlaarsdam, Van Waes and Daems (2007) found significant improvements on writing performance with university students who learned by observation, more recent studies with this population (in addition to ours) have yielded mixed results. For example, in a small-scale study with university students by Van de Weijer, Åkerlund, Johansson & Sahlén (2018), no effects of observational learning were found. In their study, they compared writing performance from students with and without a hearing impairment, who learned by observing an expert model, to the writing performance of students without a hearing impairment in a control group. Students in all groups performed equally well on argumentative writing. In a different study by Callinan, Van der Zee and Wilson (2018), which is more comparable to our studies presented in Chapters 2 and 3 no effects of observational learning were found either. In their study with 142 participants, Callinan et al. compared three groups: students observing models, students observing models and evaluating the models' performance (comparable to our observational learning with reflection condition), and students in a control group. The students in both observational learning conditions did not outperform the students in the control group on essay writing performance, nor in self efficacy beliefs. There were also no differences between the group who merely observed and the group that observed and evaluated the models' performance, indicating that reflective activities, like in our studies, did not support learning. This makes it unlikely that we did not find effects, just because university students participated in our studies. In addition, it is also worth noting that within the motor skill domain the effects of mere observation have shown to be much more varied, as we have argued in Chapter 5, and also within other areas of the writing domain, mixed results have been published: for example, Grenner et al. (2018) report only short-term effects of observational learning on argumentative writing for high school students, and Groenendijk, Janssen, Rijlaarsdam and Van den Bergh (2013) find no positive effects of observational learning on poetry writing. The findings of this dissertation, combined with the above-mentioned studies, seem to shed a more nuanced light on the effectiveness of mere observational learning, at least within the context of learning to write complex texts, suggesting they might not be as robust as was previously thought.

Learning to write is complex!

Even though this outcome might be surprising given the number of studies showing the positive effects of learning by observation, there might be good reasons for the absence of the effect in our studies. As argued earlier, writing, and especially learning to write, is a complex process from a cognitive perspective. Each writer makes use of his or her own 'cognitive architecture' (Graham, 2018). This architecture consists of long-term memory resources (such as knowledge about writing, but also prior content knowledge), control mechanisms (attention, working memory and executive control) and production processes (such as ideation, translation, and reconceptualization). However, there is arguably a lot of variability between writers in their cognitive capacities, which affects how they conceptualize a writing task and how they engage in this writing task.

This aligns with more general insights on learning. For example, Kalyuga, Ayres, Chandler and Sweller (2003) have argued that inexperienced learners, with low prior knowledge, learn better from other types of learning activities than experienced learners, with higher prior knowledge. This 'expertise reversal effect' has been found for several learning effects, such as the modality effect and the worked example effect. This last one is particularly interesting since worked examples can be linked to observational learning, as they are both types of example-based learning. With worked examples learners are presented with a written account of a problem and worked-out solution steps leading to the final solution, while in observational learning models demonstrate a problem-solving or performance procedure to a learner. It could therefore be argued that the effectiveness of observational learning to write might also be affected by learners' cognitive capacities, such as prior knowledge.

In addition, cognitive processes are not the only components affecting writing performance. Writing can also be seen as taking place within a community of writing, that is, "a group of people who share a basic set of goals and assumptions and use writing to achieve their purposes" (Graham, 2018, p. 259). Within such a community, factors such as the valued purposes of writing, and the cultural and social structure of the community, shape writing. Ng, Graham, Liu, Lau and Tang (2021) for example have studied the relations between writing motives, such as social recognition, curiosity, competition and grades, with writing self-efficacy and the time spent on tasks with 4th grade Chinese students. Their findings paint a rich but complex picture of these relations, displaying a broad range of individual differences between students in the motives they have for writing, and how these motives affect their writing motivation and the confidence they have in themselves as writers.

When we look at writing from this perspective, it can well be argued that short interventions, directed at only a limited set of factors involving writing, may not solicit robust effects, since there are many other, both cognitive and affective factors that might determine writing performance. For example, when we look at our first two studies (Chapters 2 and 3), the academic writing course in which the interventions took place can be considered the writing community with the students and the teachers as members of the community. In this community, the beliefs and values of the community, the characteristics of the writing products, as well as the subject matter were new to the students. With the variability in the students' capabilities in mind, it is conceivable that other factors than the interventions, which were mostly directed at synthesizing sources, affected writing performance. For example, in a recent, small scale study, Ten Peze, Janssen, Rijlaarsdam and Van Weijen (2021) report a relation between learner characteristics, writing processes and text quality. Students who had a positive attitude towards writing and who believed that writing requires personal commitment wrote higher quality texts, which might be explained by how much effort a writer invests in a task. However, the relations seemed to vary over different types of tasks (creative versus argumentative writing). This underlines the statement that learning to write is complex, since different, cognitive and affective, factors might affect writing in a different manner with different types of students with different types of tasks, which makes it relevant to gain more insight into these factors, since they arguably also influence the effectiveness of interventions to improve academic writing.

Designing effective learning-to-write interventions

This brings us to another aspect: the design of interventions. If writing is affected by different factors in different circumstances, how can we design effective interventions?

In this dissertation, we closely followed (and expanded) the methods and procedures of previous studies, which led to relatively short interventions. This allowed us to test the direct effects of observational learning and reflection, which led us to conclude that a short and focused intervention does not sort robust effects of observational learning and learning doing. However, due the complexity of learning to write within a new community, it could be argued that effective writing instruction should entail a more comprehensive program, when it comes to writing complex texts at university level. Such programs could for example be based on the Self-Regulated Strategy Development approach by Graham and Harris (2018) or the multifaceted approach suggested by Koster and Bouwer (2018), both including strategy instruction, self-regulation instruction, explicit instruction, gradual release of responsibility, and, observational learning. An example of an undergraduate writing program based on the SRSD approach was developed and studied by MacArthur, Philippakos and Ianetta (2015). In this semester-long program the activities were focused on developing writing strategies and self-regulated learning strategies. Observational learning, with the instructors as models, was part of the instruction, but it was embedded in a larger program, including discussing genres and their characteristics, and scaffolding. The program resulted in an increase in writing quality and self-efficacy. With an approach like this, it is not possible to determine the relative contribution of each component of the curriculum, but it does seem to do justice to the complexity of learning to write within the academic writing community.

This comprehensive approach aligns with recent developments in educational practices in which the focus is shifting more and more towards learning processes instead of merely focusing on final (writing) products (Vandermeulen, Leijten, & Van Waes, 2020). Since writing products are the result of writing processes, this seems to be a logical step, especially since students entering university have limited awareness of the writing strategies they employ and the choices they make while writing. When students enter university, they are confronted with new genres of writing: genres that are more extensive and complex than the ones they have experienced mostly during their secondary education. When engaging in executing those complex writing tasks, they usually rely on the approaches they used in high school for far less complex and shorter writing assignments. However, in most cases these ‘old’ approaches might not suffice. Changing their approach does not come easy, however. Their knowledge on effective writing strategies and the characteristics of the new text genres is mostly limited.

Observing models can be a first step in identifying writing strategies than the ones they are familiar with: even though we found no differences between observational learning and learning by doing, observational learning also did not *hinder* learning to write, which means it could still offer opportunities for writing education. In fact, this might be especially so for online or blended education, since observational learning, especially in the form of video models, can easily be integrated into online learning environments. In these online environments, it would even be possible to add interactive cues to the modeling videos, drawing attention to relevant processes which could aid learners in actively processing the learning materials (Wouters, Tabbers, & Paas, 2007).

But, as mentioned earlier, we believe it is important that observational learning is combined with other activities. For example: in observational learning the focus is mostly directed at the model's performance. Even though this may give students new ideas on how to deal with writing tasks, it does not necessarily give them insight into the actual writing strategies they themselves employ. A new, promising approach focusing on students' own writing processes is that of Vandermeulen et al. (2020). In an intervention study, they provided high school students with a process-oriented feedback report based on keystroke logging data from the students' writing sessions. The feedback report contains information on numerical data on process variables (such as the number of words, and the amount of time spent reading and writing), distribution of certain writing activities over the beginning, middle and end of the writing process, and a process graph. Students were encouraged to reflect, with their peers, on their process reports and compare them to benchmark reports. This led to a significant improvement in writing performance, comparable to the improvement usually made in one grade of schooling. Vandermeulen et al. (2020) also stress the importance of embedding the process report in a 'feedback flow' in which reflection, teacher-student and peer interaction play a vital role, which is again a call for a comprehensive approach.

Future Directions

We set out this dissertation to contribute to a growing body of evidence-based writing instruction by conceptually replicating different studies on the effects of observational learning from models. The fact that we were not able to replicate the previous findings within our own specific context, academic writing at a Dutch university, underlines the importance of replication studies when it comes the learning sciences and, more specifically research on writing instruction. The findings of our study imply that even well-established approaches to learning might still be less effective than expected when used in different contexts.

In order to design effective evidence-based writing instruction for different students in different writing communities, it is therefore relevant to try to establish which individual 'building blocks' contribute in which way under which circumstances to academic writing. One way to do so is by meta-analyses, which offer great insights in effective practices found in different contexts, with different populations at different educational levels. However, as described in the introduction, meta-analyses have several limitations, one being that non-significant findings have less chance of being published and consequently less chance of being included in a meta-analysis. Another limitation, more specifically within writing instruction research, is that the descriptions of the included instructional methods are often rather brief and not very specific (Fidalgo, Harris, & Braaksma, 2018; Bouwer & De Smedt, 2018). This makes it difficult to establish precisely *what* was taught and *how* this was done, which can complicate comparing the effects of different studies but also transferring the insights from the meta-analyses to the classroom. Recently, some initiatives have been set up to tackle this issue: Rijlaarsdam, Janssen, Rietdijk and Van Weijen (2018) and Bouwer and De Smedt (2018) have formulated recommendations for standardizing the way writing interventions are being reported with regards to the context of the interventions, the theoretical and empirical grounding, the design principles, and the description of the instructional and learning activities and materials.

This standardization might also support replication research. Replication studies, both direct and conceptual, can offer valuable insights into the generalizability of writing interventions, provided that these are described in a detailed manner. They offer opportunities to better understand which specific factors play a role under which circumstances, which can help instructors to determine whether a specific instructional component is likely to contribute to the performance of the learners in their classes. We therefore argue that replication studies should be an integrated part of the learning sciences and, more specifically, the writing research community. We believe replication studies will help us investigate the extent to which writing research findings will generalize over groups, instruments, skills, and instructions beyond the original studies, which in turn will lead to more effective writing instruction in classrooms.

Final Thoughts

In this dissertation, we investigated the role of observational learning, learning by doing and reflection in the context of academic writing. From our findings we conclude that observational learning and learning by doing, with or without reflection, result in equal academic writing performances and self-efficacy beliefs. There is an old Chinese proverb, attributed to Confucian philosopher Xunzi, that states: “Tell me and I forget, show me and I may remember, involve me and I learn.” In this dissertation we have shown that a better formulation of this ancient adage might be “Tell me, show me or involve me; it is all the same. I just might learn something, but mostly learning to write is just very complex”.

Summary

Despite increased attention for academic writing instruction, university staff in the Netherlands often complains about the (lack of) quality of students' writing. In the research presented in this dissertation, we therefore set out to investigate effective instructional methods for academic writing. We have focused on the comparison between learning by doing and observational learning in relation to academic writing performance. In addition, we investigated the role of reflection in both learning by doing and observational learning. This led to the following research questions: (1) What is the effect of observational learning on text quality and self-efficacy in the context of academic writing compared to learning by doing? (2) How does reflection affect observational learning and learning by doing in academic writing? To answer these questions, we set up four studies which were all conceptual replications, which allowed us to test the rigor and robustness of previous scientific findings with regards to observational learning.

Chapter 2 – Learning to write an academic synthesis text: observing versus doing

In our first study we compared observational learning with learning by doing in university synthesis writing. We wanted to test, in an actual first-year bachelor course on academic writing, whether observational learning would lead to higher quality texts and whether it would affect self-efficacy. The study was a conceptual replication of Raedts et al. (2007) and closely followed the procedures of that study.

In an experiment 145 undergraduate students were assigned to either an observational learning or learning by doing condition. In observational learning participants learned by observing a weak and strong models' writing processes. In learning by doing, they learned by performing writing tasks. Prior to the sessions participants were labeled as either planners or revisers based on a writing style questionnaire. The effects of the sessions were analyzed with a 2x2 between-subjects design with instructional method (observational learning, learning by doing) and writing preference (plan, revise) as factors. To measure academic writing performance the participants wrote an introduction to an empirical research paper. We found no main effects for instructional method and writing preference. Simple effect analyses did reveal that revisers benefitted somewhat more from observational learning than planners. Planners performed equally well in observational learning and learning by doing. However, planners who learned by doing did seem to outperform revisers who learned by doing.

From this chapter, we concluded that there seemed to be some positive albeit weak indications that observational learning presented interesting opportunities for learning to write, but

more work was needed. In our next study, we looked at the comparison between methods in more detail, teasing apart the effect that reflection might have. In observational learning of a cognitive skill, learners are often asked to reflect on the models' behavior explicitly and actively, while it is mostly absent in the learning by doing activities. Since reflective activities in general arguably enhance learning, it is possible that reflection mitigated the effects of observational learning.

Chapter 3 – Exploring the role of reflection in observational learning and learning by doing

In our second study, we systematically compared observational learning and learning by doing, both with and without reflection, again in the context of an academic writing course. Our goal was to investigate how reflection and instructional method affect academic synthesis writing performance, self-efficacy beliefs and students' satisfaction with the learning activities. In a quasi-experiment, 111 undergraduate students were assigned to either an observational learning or learning by doing condition, with or without reflection. Similar to the set-up in our first study, students in the observational learning condition learned by observing a weak and strong model's writing processes. In the learning by doing condition, they learned by performing writing tasks. Half of the students reflected on either the models' or their own performance. We found that reflection did not affect academic writing performance and self-efficacy beliefs, and neither did instructional method. Both reflection and instructional method did influence students' satisfaction with the learning activities. Students preferred learning by doing over observational learning, and reflecting over not reflecting. From this study, we concluded that in academic synthesis writing the interplay between reflection, observational learning and learning by doing was not evident yet: students seemed to perform equally well in all conditions, even though they preferred learning by doing over observational learning, and reflecting over not reflecting.

The two single subject design studies from Chapters 2 and 3 were conducted in a 'real life' writing course. This supported the ecological validity of the studies, but it complicated controlling for other factors that might be of influence. In our next study, we therefore attempted to study observational learning and reflection in a more controlled environment (a lab study) and with a less complex writing task (sentence combining).

Chapter 4 – Combining sentences: comparing instructional methods with a less complex writing task

In our first two studies, we were not able to replicate previous findings on observational learning with a complex synthesis writing task. In our third study, we therefore investigated the effects of instructional method and reflection in a more controlled environment and a less complex writing task, namely sentence combining, by conceptually replicating Zimmerman and Kitsantas (2002). 108 undergraduate students were taught sentence combining by either watching a coping model or a mastery model performing sentence combining tasks, or by practicing with sentence combining tasks themselves. Additionally, half of the students in either instructional method was explicitly asked to reflect on the learning process, while the other half was not. From our findings, we concluded that neither modeling type nor reflection affected writing revision proficiency. Students who observed a model did show an initial increase in their self-efficacy beliefs which disappeared after a practice round. Learning from both the coping and mastery model did lead to a more accurate implementation of the proposed solution strategy and the students preferred observing a model over mere practicing. This did however not lead to higher performance scores.

In Chapters 2, 3 and 4, we focused on the effectiveness of observational learning and learning by doing, with and without reflective activities in the cognitive, writing domain. In our next study, we shifted our focus to motor skill acquisition, to test to what extent our findings from a cognitive domain would transpose to the motor skill domain.

Chapter 5 – Observational learning and learning by doing in motor skill acquisition

In the last study of this dissertation, we investigated the effects of modeling type (observational learning from an expert model or a novice model versus learning by doing) and reflection on the acquisition of dart-throwing skills, self-efficacy beliefs and self-reaction scores by conceptually replicating a study by Kitsantas, Zimmerman, and Cleary (2000). Based on Kitsantas et al., participants observing a novice model were expected to surpass participants observing an expert model who in turn were expected to outperform participants who learned without a model. Reflection was hypothesized to have a positive effect. 156 High school and university students were tested three times: in a pre-test, after a modeling intervention, and after a practice round. Again, we found no main effects of modeling type and reflection. No interaction effects were found either. There was an effect of testing moment, indicating that participants improved dart-throwing skills, self-efficacy beliefs, and self-reaction scores over time. With these findings, we were not able to replicate Kitsantas et al. From our study, we concluded that observational learning, irrespective of the model's skill level, combined with physical practice, yields similar results as mere physical practice.

General Conclusion

In this dissertation, we investigated the role of observational learning, learning by doing and reflection in the context of academic writing. From our findings we conclude that observational learning and learning by doing, with or without reflection, result in equal academic writing performances and self-efficacy beliefs. We find this consistently with different types of tasks and in different contexts. The fact that we were not able to replicate previous findings within our own specific context, academic writing at a Dutch university, underlines the importance of replication studies when it comes the learning sciences and, more specifically research on writing instruction. The findings of our study imply that even well-established approaches to learning might still be less effective than expected when used in different contexts.

References

Aarts, R., de Jong, G., Madern, A., Diephuis, E., & van der Kuip, J. (2019). *Academische vaardigheden in voortgezet onderwijs en aansluiting bij het hoger onderwijs; een inventarisatie: In het kader van het Project Academische Vaardigheden, Tilburg University. Deel 1 2018-2019.*

Andrews, R., Torgerson, C., Beverton, S., Freeman, A., Locke, T., Low, G., ... & Zhu, D. (2006). The effect of grammar teaching on writing development. *British Educational Research Journal, 32*(1), 39-55. https://doi.org/10.1080/01411920500401997

Applebee, A. N. (1984). Writing and reasoning. *Review of Educational Research, 54*(4), 577-596. https://doi.org/10.3102/00346543054004577

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191-215

Bandura, A. (1977). *Social learning theory.* Englewood Cliffs, N.J.: Prentice Hall.

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

Bandura, A. (2016). The Power of Observational Learning Through Social Modeling. In R.J. Sternberg, S.T. Fiske & D.J. Foss (Eds.), S*cientists Making a Difference: One Hundred Eminent Behavioral and Brain Scientists Talk about Their Most Important Contributions* (pp. 235-239). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781316422250.052

Blanchard, A. (2014). *Bull's eye-hand coordination: visual and motor contributions to observational learning* (Doctoral dissertation, Rutgers University-Graduate School-Newark).

Blandin, Y., Lhuisset, L., & Proteau, L. (1999). Cognitive processes underlying observational learning of motor skills. *The Quarterly Journal of Experimental Psychology: Section A, 52*(4), 957-979. https://doi.org/10.1080/713755856

Blandin, Y., Proteau, L., & Alain, C. (1994). On the cognitive processes underlying contextual interference and observational learning. *Journal of Motor Behavior, 26*(1), 18-26. https://doi.org/10.1080/00222895.1994.9941657

Bonset, H. (2010). Nederlands in het voortgezet en hoger onderwijs: Hoe sluit dat aan? Deel 1. *Levende Talen Magazine*, *97*(3), 16-20.

Boscolo, P., Arfé, B., & Quarisa, M. (2007). Improving the quality of students' academic writing: an intervention study*. Studies in Higher Education, 32*(4), 419-438. https://doi.org/10.1080/03075070701476092

Boud, D., Keogh, R., & Walker, D. (2005). *Reflection: Turning Experience into Learning.* London: RoutledgeFalmer.

Bouwer, R., & De Smedt, F. (2018). Introduction Special Issue: Considerations and Recommendations for Reporting on Writing Interventions in Research Publications. *Journal of Writing Research, 10(2),* 115-137. doi: 10.17239/jowr-2018.10.01.01

Braaksma, M. A. H. (2002). *Observational learning in argumentative writing* (Doctoral Dissertation). Amsterdam: Graduate School for Teaching and Learning, University of Amsterdam.

Braaksma, M. A., Rijlaarsdam, G., & Van den Bergh, H. (2002). Observational learning and the effects of model-observer similarity. *Journal of Educational Psychology, 94*(2), 405. https://doi.org/10.1037//0022-0663.94.2.405

Braaksma, M. A., Rijlaarsdam, G., Van den Bergh, H., & van Hout-Wolters, B. H. M. (2004). Observational learning and its effects on the orchestration of writing processes. *Cognition and Instruction, 22*(1), 1-36. https://doi.org/10.1207/s1532690xci2201_1

Braaksma, M. A., van den Bergh, H., Rijlaarsdam, G., & Couzijn, M. (2001). Effective learning activities in observation taks when learning to write and read argumentative texts. *European Journal of Psychology of Education, 16*(1), 33. https://doi.org/10.1007/BF03172993

Braaksma, M. A., 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, 16*(1), 33.

Braaksma, M.A.H., Rijlaarsdam, G., Van den Bergh, H., & Van Hout-Wolters, B.H.A.M. (2006). What observational learning in writing courses entails: A multiple case study. *L1-Educational Studies in Language and Literature, 6*(1), pp. 31-62. https://doi.org/10.17239/L1ESLL-2006.06.01.05

Broadbent, J., & Poon, W. L. (2015). Self-regulated learning strategies & academic achievement in online higher education learning environments: A systematic review. *The Internet and Higher Education, 27*, 1-13. https://doi.org/10.1016/j.iheduc.2015.04.007

Bruning, R., & Kauffman, D. F. (2016). Self-efficacy beliefs and motivation in writing development. In C. MacArthur, S. Graham, & J. Fitzgerald (Eds.), Handbook of writing research (2nd ed., pp. 96–114). New York, NY: Guilford Press.

Bruning, R., Dempsey, M., Kauffman, D. F., McKim, C., & Zumbrunn, S. (2013). Examining dimensions of self-efficacy for writing. *Journal of Educational Psychology, 105*(1), 25-38. https://doi.org/10.1037/a0029692

Callinan, C. J., van der Zee, E., & Wilson, G. (2018). Developing essay writing skills: an evaluation of the modelling behaviour method and the influence of student self-efficacy. *Journal of Further and Higher Education, 42*(5), 608-622. https://doi.org/10.1080/0309877X.2017.1302564

Carroll, W. R., & Bandura, A. (1987). Translating cognition into action: The role of visual guidance in observational learning. *Journal of Motor Behavior, 19*(3), 385-398. https://doi.org/10.1080/00222895.1987.10735419

Carter, B. J. (2010). Evidence-based decision-making: practical issues in the appraisal of evidence to inform policy and practice. Australian Health Review, 34(4), 435-440.Davies, P. (1999). What is evidence-based education? *British Journal of Educational Studies*, *47*(2), 108-121.

Chang, M. M. (2007). Enhancing web-based language learning through self-monitoring. *Journal of Computer Assisted Learning, 23*(3), 187-196. https://doi.org/10.1111/j.1365-2729.2006.00203.x

Clark, R. E. (1982). Antagonism between achievement and enjoyment in ATI studies. *Educational Psychologist, 17*(2), 92-101. https://doi.org/10.1080/00461528209529247

Cordovani, L., & Cordovani, D. (2016). A literature review on observational learning for medical motor skills and anesthesia teaching. *Advances in Health Sciences Education, 21*(5), 1113-1121. https://doi.org/10.1007/s10459-015-9646-5

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. https://doi.org/10.1016/s0959-4752(98)00040-1

Credé, M., Roch, S. G., & Kieszczynka, U. M. (2010). Class attendance in college: A meta-analytic review of the relationship of class attendance with grades and student characteristics. *Review of Educational Research*, *80*(2), 272-295. DOI: 10.3102/0034654310362998

D'Innocenzo, G., Gonzalez, C. C., Williams, A. M., & Bishop, D. T. (2016). Looking to learn: the effects of visual guidance on observational learning of the golf swing. *PloS one, 11*(5), e0155442. https://doi.org/10.1371/journal.pone.0155442

Darden, G. F. (1997). Demonstrating motor skills—rethinking that expert demonstration. *Journal of Physical Education, Recreation & Dance, 68*(6), 31-35. https://doi.org/10.1080/07303084.1997.10604962

Daudelin, M. W. (1996). Learning from experience through reflection. *Organizational Dynamics, 24(*3), 36-48. https://doi.org/10.1016/S0090-2616(96)90004-2

De Smet, M. J., Brand-Gruwel, S., Leijten, M., & Kirschner, P. A. (2014). Electronic outlining as a writing strategy: Effects on students' writing products, mental effort and writing process. *Computers & Education, 78,* 352-366. https://doi.org/10.1016/j.compedu.2014.06.010

Dempsey, M. S., & Kauffman, D. F. (2017). Supporting Third Year Medical Students' Skill Acquisition and Self-Efficacy with Coping Models and Process Feedback during Laparoscopic Knot Tying Simulation. *Frontiers in Psychology, 8*, 1171. https://doi.org/10.3389/fpsyg.2017.01171

Dörnyei, Z. (2009). Individual differences: Interplay of learner characteristics and learning environment. *Language Learning, 59*, 230-248. https://doi.org/10.1111/j.1467-9922.2009.00542.x

Elbow, P. (1998). Writing without teachers (2nd edition). Oxford: Oxford University Press

Ferrari, M. (1996). Observing the observer: Self-regulation in the observational learning of motor skills. *Developmental Review*, *16*(2), 203-240. https://doi.org/10.1006/drev.1996.0008

Fidalgo, R., Harris, K., & Braaksma, M. (2018). *Design Principles for Teaching Effective Writing* (Vol. 34). Leiden; Boston: Brill.

Fidalgo, R., Torrance, M., Rijlaarsdam, G., van den Bergh, H., & Álvarez, M. L. (2015). Strategy-focused writing instruction: Just observing and reflecting on a model benefits 6th grade students. *Contemporary Educational Psychology, 41*, 37-50. https://doi.org/10.1016/j.cedpsych.2014.11.004

Fischman, M. G. (2015). On the continuing problem of inappropriate learning measures: Comment on Wulf et al. (2014) and Wulf et al. (2015). *Human Movement Science, 42*, 225-231. https://doi.org/10.1016/j.humov.2015.05.011

Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication,* 365-387. https://doi.org/10.2307/356600

Froese, A. D., Gantz, B. S., & Henry, A. L. (1998). Teaching students to write literature reviews: A meta-analytic model. *Teaching of Psychology*, 25, 102-105. https://doi.org/10.1207/s15328023top2502_4

Galbraith, D., & Torrance, M. (2004). Revision in the Context of Different Drafting Strategies. In: L. Allal, L. Chanquoy, & P. Largy (Eds.), *Revision Cognitive and Instructional Processes. Studies in Writing*, *vol 13*. Springer, Dordrecht. https://doi.org/10.1007/978-94-007-1048-1_5

Galbraith, D., Torrance, M., & Hallam, J. (2006). Effects of writing on conceptual coherence. In Proceedings of the 28th annual conference of the Cognitive Science Society (pp. 1340-1345).

Gogol, K., Brunner, M., Martin, R., Preckel, F., & Götz, T. (2017). Affect and motivation within and between school subjects: Development and validation of an integrative structural model of academic self-concept, interest, and anxiety. *Contemporary Educational Psychology, 49*, 46-65. https://doi.org/10.1016/j.cedpsych.2016.11.003

Graham, S. (2018). A writer(s) within community model of writing. In C. Bazerman, V. Berninger, D. Brandt, S. Graham, J. Langer, S. Murphy, P. Matsuda, D. Rowe, & M. Schleppegrell (Eds.), *The lifespan development of writing* (pp. 271-325). Urbana, IL: National Council of English.

Graham, S. (2019). Changing how writing is taught. *Review of Research in Education, 43*(1), 277-303. https://doi.org/10.3102%2F0091732X18821125

Graham, S., & Harris, K. (2018b). An examination of the design principles underlying a Self-Regulated Strategy Development study. *Journal of Writing Research, 10*(2), 139-187. doi:10.17239/jowr-2018.10.01.02

Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology, 99*(3), 445-476. https://doi.org/10.1037/0022-0663.99.3.445

Graham, S., Gillespie, A., & McKeown, D. (2013). Writing: Importance, development, and instruction. *Reading and Writing, 26*(1), 1-15. doi: 10.1007/s11145-012-9395-2

Graham, S., Harris, K. R., & Chambers, A. B. (2016). Evidence-based practice and writing instruction. *Handbook of Writing Research, 2*, 211-226.

Graham, S., Harris, K. R., & Mason, L. (2005). Improving the writing performance, knowledge, and self-efficacy of struggling young writers: The effects of self-regulated strategy development. *Contemporary Educational Psychology*, *30*(2), 207-241.

Graham, S., McKeown, D., Kiuhara, S., & Harris, K. R. (2012). A meta-analysis of writing instruction for students in the elementary grades. *Journal of Educational Psychology, 104*(4), 879-896. http://dx.doi.org/10.1037/a0029185

Granello, D. H. (2001). Promoting cognitive complexity in graduate written work: Using Bloom's taxonomy as a pedagogical tool to improve literature reviews. *Counselor Education and Supervision, 40*(4), 292-307. https://doi.org/10.1002/j.1556-6978.2001.tb01261.x

Green, R., & Bowser, M. (2006). Observations from the field: Sharing a literature review rubric. *Journal of Library Administration, 45*(1-2), 185-202. https://doi.org/10.1300/j111v45n01_10

Grenner, E., Åkerlund, V., Asker-Árnason, L., van de Weijer, J., Johansson, V., & Sahlén, B. (2020). Improving narrative writing skills through observational learning: a study of Swedish 5th-grade students. *Educational Review, 72*(6), 691-710. https://doi.org/10.1080/00131911.2018.1536035

Groenendijk, T., Janssen, T., Rijlaarsdam, G., & van den Bergh, H. (2013). Learning to be creative. The effects of observational learning on students' design products and processes. *Learning and Instruction, 28*, 35-47. https://doi.org/10.1016/j.learninstruc.2013.05.001

Groenendijk, T., Janssen, T., Rijlaarsdam, G., & van den Bergh, H. (2013). The effect of observational learning on students' performance, processes, and motivation in two creative domains. *British Journal of Educational Psychology*, 83(1), 3-28. https://doi.org/10.1111/j.2044-8279.2011.02052.x

Hancock, G. R., Butler, M. S., & Fischman, M. G. (1995). On the problem of two-dimensional error scores: Measures and analyses of accuracy, bias, and consistency. *Journal of Motor Behavior, 27*(3), 241-250. https://doi.org/10.1080/00222895.1995.9941714

Hjørland, B. (2011). Evidence-based practice: An analysis based on the philosophy of science. *Journal of the American Society for Information Science and Technology, 62*(7), 1301-1310. https://doi.org/10.1002/asi.21523

Horn, R. R., Williams, A. M., Scott, M. A., & Hodges, N. J. (2005). Visual search and coordination changes in response to video and point-light demonstrations without KR. *Journal of Motor Behavior*, *37*(4), 265-274.

Janelle, C. M., & Hillman, C. H. (2003). Expert performance in sport: Current perspectives and critical issues. In J. L. Starkes &K. A. Ericsson (Eds.), *Expert performance in sports: Advances in research on sport expertise* (pp. 49–83). Champaign, IL: Human Kinetics.

Kalyuga, Ayres, Chandler, & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 38(*1), 23–31.

Kellogg, R.T. (2008). Training writing skills: A cognitive developmental perspective. *Journal of Writing Research, 1*(1), 1-26. https://doi.org/10.17239/jowr-2008.01.01.1

Kieft, M., Rijlaarsdam, G., & van den Bergh, H. (2006). Writing as a learning tool: testing the role of students' writing strategies. *European Journal of Psychology of Education, 21*(1), 17-34. https://doi.org/10.1007/bf03173567

Kieft, M., Rijlaarsdam, G., & van den Bergh, H. (2008). An aptitude–treatment interaction approach to writing-to-learn. *Learning and Instruction, 18*(4), 379-390. https://doi.org/10.1016/j.learninstruc.2007.07.004

Kirschner, P. A. (2017). Stop propagating the learning styles myth. *Computers & Education*, 106, 166-171. https://doi.org/10.1016/j.compedu.2016.12.006

Kirschner, P. A., & van Merriënboer, J. J. (2013). Do learners really know best? Urban legends in education. *Educational psychologist, 48*(3), 169-183. https://doi.org/10.1080/00461520.2013.804395

Kitsantas, A., Zimmerman, B. J., & Cleary, T. (2000). The role of observation and emulation in the development of athletic self-regulation. *Journal of Educational Psychology*, *92*(4), 811. https://doi.org/10.1037/0022-0663.92.4.811

Koster, J. (1975). Dutch as an SOV language. *Linguistic Analysis, 1*(2), 111-136.

Koster, M. P., & Bouwer, I. R. (2016). *Bringing writing research into the classroom: The effectiveness of Tekster, a newly developed writing program for elementary students* (Doctoral dissertation, University Utrecht). Utrecht: Interuniversity Centre for Educational Research.

Koster, M., & Bouwer, R. (2018). Describing multifaceted writing interventions: From design principles for the focus and mode of instruction to student and teacher activities. *Journal of Writing Research, 10*(2), 189-224. doi: 10.17239/jowr-2018.10.01.03

Lakens, D. (2017). Equivalence tests: a practical primer for t tests, correlations, and meta-analyses. *Social Psychological and Personality Science, 8*(4), 355-362. https://doi.org/10.1177/1948550617697177

Law, B., & Hall, C. (2009). Observational learning use and self-efficacy beliefs in adult sport novices. *Psychology of sport and exercise, 10*(2), 263-270. https://doi.org/10.1016/j.psychsport.2008.08.003

Lee, T. D., & White, M. A. (1990). Influence of an unskilled model's practice schedule on observational motor learning. *Human Movement Science*, *9*(3), 349-367. https://doi.org/10.1016/0167-9457(90)90008-2

Limpo, T., & Alves, R. A. (2013). Teaching planning or sentence-combining strategies: Effective SRSD interventions at different levels of written composition. *Contemporary Educational Psychology, 38*(4), 328-341. https://doi.org/10.1016/j.cedpsych.2013.07.004

Lynch Jr, J. G., Bradlow, E. T., Huber, J. C., & Lehmann, D. R. (2015). Reflections on the replication corner: In praise of conceptual replications. *International Journal of Research in Marketing*, *32*(4), 333-342. https://doi.org/10.1016/j.ijresmar.2015.09.006

MacArthur, C. A., Philippakos, Z. A., & Ianetta, M. (2015). Self-regulated strategy instruction in college developmental writing. *Journal of Educational Psychology, 107*(3), 855-867. https://doi.org/10.1037/edu0000011

Makel, M. C., & Plucker, J. A. (2014). Facts are more important than novelty: Replication in the education sciences. *Educational Research, 20*(10), 1-13. https://doi.org/10.3102/0013189X14545513

Makel, M. C., Smith, K. N., McBee, M., Peters, S. J., & Miller, E. M. (2019, March 26). Open Science 2.0: Large-Scale Collaborative Education Research. *PsyArXiv* https://doi.org/10.31234/osf.io/ypmjg

Marks, D. F. & Yardley, L. (2004). Synthesising evidence: systematic reviews, meta-analysis and preference analysis. In Marks, D. F., & Yardley, L. *Research Methods for Clinical and Health Psychology* (pp. 185-209). London: SAGE Publications, Ltd. doi: 10.4135/9781849209793

Mateos, M., & Solé, I. (2009). Synthesising information from various texts: A study of procedures and products at different educational levels. *European Journal of Psychology of Education, 24*(4), 435. https://doi.org/10.1007/BF03178760

McNamara, D. S., & Kintsch, W. (1996). Learning from texts: Effects of prior knowledge and text coherence. *Discourse processes, 22*(3), 247-288. https://doi.org/10.1080/01638539609544975

Mega, C., Ronconi, L., & De Beni, R. (2014). What makes a good student? How emotions, self-regulated learning, and motivation contribute to academic achievement. *Journal of Educational Psychology, 106*(1), 121-131. https://doi.org/10.1037/a0033546

Monfardini, E., Gazzola, V., Boussaoud, D., Brovelli, A., Keysers, C., & Wicker, B. (2013). Vicarious neural processing of outcomes during observational learning. *PloS one, 8*(9), e73879. https://doi.org/10.1371/journal.pone.0073879

Moore, C. M., Lelievre, N., & Ste-Marie, D. M. (2019). Observing different model types interspersed with physical practice has no effect on consolidation or motor learning of an elbow flexion–extension task. *Human Movement Science, 63*, 96-107. https://doi.org/10.1016/j.humov.2018.11.014

Murray, D. M. (1978). Internal revision: A process of discovery. In C. R. Cooper & L. Odell (Eds.), *Research on composing: Points of departure* (pp. 85-103). Urbana, IL: National Council of Teachers of English.

Nelson, J., & Campbell, C. (2017). Evidence-informed practice in education: meanings and applications. *Educational Research*, *59*(2), 127-135.

Nosek, B. A., & Errington, T. M. (2017). Reproducibility in cancer biology: Making sense of replications. *Elife*, 6, e23383. https://doi.org/10.7554/eLife.23383.001

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), aac4716. http://doi.org/10.1126/science.aac4716

Pollock, B. J., & Lee, T. D. (1992). Effects of the model's skill level on observational motor learning. *Research quarterly for exercise and sport, 63*(1), 25-29. https://doi.org/10.1080/02701367.1992.10607553

Raedts, M., Daems, F., Van Waes, L., & Rijlaarsdam, G. (2009). Observational learning through peer models in a complex writing task. *Tijdschrift voor Taalbeheersing, 31*(9), 142-165. https://doi.org/10.5117/tvt2009.2.obse357

Raedts, M., Rijlaarsdam, G., Van Waes, L., & Daems, F. (2007). Observational learning through video-based models: impact on student's accuracy of self-efficacy beliefs, task knowledge and writing performances. In G. Rijlaarsdam (Series Ed.) and P. Boscolo & S. Hidi (Volume Eds.), *Studies in Writing, Volume 19, Writing and Motivation* (pp. 219–238). Oxford: Elsevier. https://doi.org/10.1163/9781849508216_013

Raedts, M., Van Steendam, E., De Grez, L., Hendrickx, J., & Masui, C. (2017). The effect of different types of video modelling on undergraduate students' motivation and learning in an academic writing course. *Journal of Writing Research, 8*(3), 399-435. https://doi.org/10.17239/jowr-2017.08.03.01

Rak, N., Bellebaum, C., & Thoma, P. (2013). Empathy and feedback processing in active and observational learning. *Cognitive, Affective, & Behavioral Neuroscience*, *13*(4), 869-884. https://doi.org/10.3758/s13415-013-0187-1

Ram, N., Riggs, S. M., Skaling, S., Landers, D. M., & McCullagh, P. (2007). A comparison of modelling and imagery in the acquisition and retention of motor skills. *Journal of Sports Sciences, 25*(5), 587-597. https://doi.org/10.1080/02640410600947132

Rijlaarsdam, G., & Couzijn, M. (2000). Writing and learning to write: A double challenge. In R. Simons, J. Van der Linden, & T. Duffy (Eds.), *New Learning* (pp. 157-189). Dordrecht: Kluwer. https://doi.org/10.1007/0-306-47614-2_9

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. https://doi.org/10.17239/jowr-2008.01.01.3

Rijlaarsdam, G., Janssen, T., Rietdijk, S., & van Weijen, D. (2018). Reporting design principles for effective instruction of writing: Interventions as constructs. In R. Fidalgo, K. Harris, & M. Braaksma (Eds.), *Design Principles for Teaching Effective Writing* (Vol. 34, pp. 280-313). Leiden; Boston: Brill.

Rohbanfard, H., & Proteau, L. (2011). Learning through observation: a combination of expert and novice models favors learning. *Experimental Brain Research, 215*(3-4), 183-197. https://doi.org/10.1007/s00221-011-2882-x

Saddler, B., & Graham, S. (2005). The effects of peer-assisted sentence combining instruction on the writing performance of more and less skilled young writers. *Journal of Educational Psychology, 97*(1), 43–54. https://psycnet.apa.org/doi/10.1037/0022-0663.97.1.43

Scardamalia, M., & Bereiter, C. (1987). Knowledge telling and knowledge transforming in written composition. In S. Rosenberg (Ed.), *Advances in Applied Psycholinguistics, 2* (pp. 308-324). Cambridge United Kingdom: Cambridge University Press.

Scargle, J. D. (2000). Publication Bias: The "File-Drawer" Problem in Scientific Inference. *Journal of Scientific Exploration*, *14*(1), 91-106. https://arxiv.org/abs/physics/9909033

Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology*, *13*(2), 90. doi: 10.1037/a0015108

Schunk D.H., Pintrich P., & Meece J. (2010). *Motivation in Education: Theory, Research, and Application Third Edition*. New Jersey: Pearson Education, Inc.

Schunk, D. H. (1983). Ability versus effort attributional feedback: Differential effects on self-efficacy and achievement. *Journal of Educational Psychology, 75*, 848-856. https://doi.org/10.1037/0022-0663.75.6.848

Schunk, D. H. (1987). Peer models and children's behavioral change. Review of Educational Research, 57(2), 149-174. https://doi.org/10.3102/00346543057002149

Schunk, D. H., & Hanson, A. R. (1985). Peer models: Influence on children's self-efficacy and achievement. *Journal of Educational Psychology, 77*, 313–322. https://doi.org/10.1037/0022-0663.77.3.313

Schunk, D. H., & Hanson, A. R. (1989). Influence of peer-model attributes on children's beliefs and learning. *Journal of Educational Psychology, 81*, 431-434. https://doi.org/10.1037/0022-0663.81.3.431

Schunk, D. H., Hanson, A. R., & Cox, P. D. (1987). Peer-model attributes and children's achievement behaviors. *Journal of Educational Psychology, 79*, 54–61.

Shea, C. H., Wright, D. L., Wulf, G., & Whitacre, C. (2000). Physical and observational practice afford unique learning opportunities. *Journal of Motor Behavior, 32*(1), 27-36. https://doi.org/10.1080/00222890009601357

Silliman, E. R., Bahr, R. H., & Wilkinson, L. C. (2020). Writing across the academic languages: introduction. *Reading and Writing, 33*(1), 1-11. https://doi.org/10.1007/s11145-019-09993-0

Simons, D. J. (2014). The Value of Direct Replication. *Perspectives on Psychological Science, 9*(1), 76–80. https://doi.org/10.1177/1745691613514755

Simonsohn, U. (2015). Small telescopes: Detectability and the evaluation of replication results. *Psychological science*, *26*(5), 559-569. https://doi.org/10.1177/0956797614567341

Sitzmann, T., & Ely, K. (2011). A meta-analysis of self-regulated learning in work-related training and educational attainment: What we know and where we need to go. *Psychological Bulletin, 137(*3), 421-442. http://dx.doi.org/10.1037/a0022777

Ste-Marie, D. M., Law, B., Rymal, A. M., Jenny, O., Hall, C., & McCullagh, P. (2012). Observation interventions for motor skill learning and performance: an applied model for the use of observation. *International Review of Sport and Exercise Psychology, 5*(2), 145-176. https://doi.org/10.1080/1750984X.2012.665076

Sweller, J., Van Merrienboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*(3), 251-296. https://doi.org/10.1023/A:1022193728205

Ten Peze, A., Janssen, T., Rijlaarsdam, G., & Van Weijen, D. (2021). Writing creative and argu-mentative texts: What's the difference? Exploring how task type affects students' writing be- haviour and performance. *L1-Educational Studies in Language and Literature, 21*, 1-38. https://doi.org/10.17239/L1ESLL-2021.21.01.11

Tillema, M. (2012). *Writing in first and second language: Empirical studies on text quality and writing processes* (Doctoral Dissertation). Utrecht: Netherlands Graduate School of Linguistics.

Torrance, M., Thomas, G. V., & Robinson, E. J. (2000). Individual differences in undergraduate essay-writing strategies: A longitudinal study. *Higher Education, 39*(2), 181-200. https://doi.org/10.1023/a:1003990432398

Van de Weijer, J., Åkerlund, V., Johansson, V., & Sahlén, B. (2019). Writing intervention in university students with normal hearing and in those with hearing impairment: can observational learning improve argumentative text writing? *Logopedics Phoniatrics Vocology, 44*(3), 115-123. https://doi.org/10.1080/14015439.2017.1418427

Van Gog, T., & Rummel, N. (2010). Example-based learning: Integrating cognitive and social-cognitive research perspectives. *Educational Psychology Review, 22*(2), 155-174. doi: 10.1007/s10648-010-9134-7

Van Steendam, E., Rijlaarsdam, G., Sercu, L., & Van den Bergh, H. (2010). The effect of instruction type and dyadic or individual emulation on the quality of higher-order peer feedback in EFL. *Learning and Instruction, 20*(4), 316-327. https://doi.org/10.1016/j.learninstruc.2009.08.009

Van Weijen, D. (2009). *Writing processes, text quality, and task effects: Empirical studies in first and second language writing* (Doctoral Dissertation). Utrecht: Netherlands Graduate School of Linguistics.

Vandermeulen, N., Leijten, M., & Van Waes, L. (2020). Reporting Writing Process Feedback in the Classroom Using Keystroke Logging Data to Reflect on Writing Processes. *Journal of Writing Research, 12*(1), 109-139. https://doi.org/https://doi.org/10.17239/jowr-2020.12.01.05

Weeks, D. L., & Anderson, L. P. (2000). The interaction of observational learning with overt practice: Effects on motor skill learning. *Acta Psychologica*, *104*(2), 259-271. https://doi.org/10.1016/S0001-6918(00)00039-1

Weir, P. L., & Leavitt, J. L. (1990). Effects of model's skill level and model's knowledge of results on the performance of a dart throwing task. *Human Movement Science, 9*(3-5), 369-383. https://doi.org/10.1016/0167-9457(90)90009-3

Wesch, N. N., Law, B., & Hall, C. R. (2007). The use of observational learning by athletes. *Journal of Sport Behavior*, *30*(2), 219.

Wouters, P., Paas, F., & van Merriënboer, J. J. (2009). Observational learning from animated models: Effects of modality and reflection on transfer. Contemporary Educational Psychology, 34(1), 1-8. https://doi.org/10.1016/j.cedpsych.2008.03.001

Wouters, P., Tabbers, H. K., & Paas, F. (2007). Interactivity in video-based models. *Educational Psychology Review, 19*(3), 327-342. DOI 10.1007/s10648-007-9045-4

Wulf, G., Shea, C., & Lewthwaite, R. (2010). Motor skill learning and performance: a review of influential factors. *Medical Education, 44*(1), 75-84. https://doi.org/10.1111/j.1365-2923.2009.03421.x

Yang, M., Badger, R., & Yu, Z. (2006). A comparative study of peer and teacher feedback in a Chinese EFL writing class. *Journal of Second Language Writing, 15*(3), 179-200. https://doi.org/10.1016/j.jslw.2006.09.004

Zimmerman, B. J. (1989). A social cognitive view of self-regulated academic learning. *Journal of Educational Psychology, 81*(3), 329-339. https://doi.org/10.1037/0022-0663.81.3.329

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal*, *45*(1), 166-183. DOI: 10.3102/0002831207312909

Zimmerman, B. J., & Kitsantas, A. (2002). Acquiring writing revision and self-regulatory skill through observation and emulation. *Journal of Educational Psychology, 94*(4), 660-668. https://doi.org/10.1037//0022-0663.94.4.660

Zimmerman, B. J., Schunk, D. H., & DiBenedetto, M. K. (2015). A personal agency view of self-regulated learning: The role of goal setting. In F. Guay, H. Marsh, D. M. McInerney, & R. G. Craven (Eds.), *International advances in self research. Self-concept, motivation and identity: Underpinning success with research and practice* (pp. 83-114). Charlotte, NC, US: IAP Information Age Publishing.

Zwaan, R. A., Etz, A., Lucas, R. E., & Donnellan, M. B. (2018). Making replication mainstream. *Behavioral and Brain Sciences, 41*(120), 1-13. doi:10.1017/S0140525X17001972

Acknowledgments

Schrijven is moeilijk! Dat is een van de conclusies van mijn onderzoek. Dat geldt natuurlijk voor academisch schrijven, maar misschien nog wel meer voor het schrijven van een dankwoord. Hoe bedank je al die mensen die elk op hun eigen manier belangrijk zijn geweest de afgelopen jaren…? Hieronder mijn beste poging:

In 2011 maakte ik een comeback op Tilburg University, als docent Academisch Nederlands. Een van mijn eerste functioneringsgesprekken verliep ongeveer zo:

EK: Hoe denk jij eigenlijk over onderzoek doen, Janneke?
JL: Eh, ja, lijkt me heel interessant. Maar ik wil liever geen ‘grote’ dingen doen.
EK: Nou, dan ga je iets kleins doen.
JL: En liefst iets praktisch.
EK: Prima.
JL: En gelinkt aan mijn eigen onderwijs!
EK: Ook prima. En je weet het, hè, Janneke: vier kleine dingen zijn samen één groot ding.

Het duurde even voordat ik me realiseerde dat met dit gesprek mijn promotietraject was gestart. Wat een kans. En ‘dat grote ding’? Dat ligt nu voor jullie.

Emiel, door dat korte gesprek werd ik volledig uit mijn comfortzone getrokken. Iets wat ik heel erg spannend vond, maar tegelijkertijd stiekem ook wel graag wilde en misschien zelfs wel nodig had. Je hebt me het hele traject lang veel vertrouwen gegeven: jij zag al veel eerder dan ik dat het heus ooit goed zou komen. Door jouw manier van begeleiden voelde ik me altijd *in the lead* terwijl ik ook de zekerheid had dat jij er was om me bij te sturen, te motiveren of mij er even aan te herinneren wat prioriteit had en wat niet. Dat laatste was zeker niet mijn sterkste kant: alles is namelijk interessant, urgent en belangrijk! Soms troffen we elkaar op BKS of Lowlands, waar je met Daan setlists stond te vangen of mij van belangrijke muzikale tips voorzag (Big Thief!). En gelukkig weet je inmiddels dat het echt niet meer cool is om je festivalbandje nog een week om te houden. Emiel, je was onmisbaar in dit traject. Dank je wel!

Marije, jij bent behalve co-promotor nog zoveel meer voor mij geweest de afgelopen jaren: collega, sparringpartner, vriendin. Dank je wel voor alles: het meedenken, je positieve en constructieve manier van feedback, het stoom afblazen (hoe lang hielden we onze anti-klaagdeal vol? 5 minuten?), het kletsen over de dingen die ons bezighielden: mijn onderzoek, de ULT, onze mannen/kinderen. Mijn hart maakte een sprongetje toen ik je voor het eerst weer hoorde fluiten in de gang van Dante. Dat heb ik zo gemist!

Ik wil ook de leden van mijn commissie heel erg bedanken. Ik bewonder uw werk enorm en ben er daarom heel trots op dat u aan deze commissie hebt willen deelnemen. Dat het een Belgisch-Nederlandse alliantie is geworden, maakt het voor mij als Nederbelg extra bijzonder.

Ad, ik hoop dat mijn verdediging niet gaat tegenvallen, na die legendarische diploma-uitreiking in 2003: de familie heeft instructies gekregen niét te zingen deze keer.

En dan mijn CIW-collega's. Zelfs tijdens Corona bleken we een ijzersterke club te zijn, mede dankzij de CIW Business en Leisure apps. Altijd goed voor praktische en morele ondersteuning, en met nerdy statistiek-humor op z'n tijd. Een speciaal woord van dank voor Naomi en Karin (Team PSI/ACNL forever!). Naomi, we zitten zó op één lijn wat betreft onze visie op onderwijs voor eerstejaars, en je bent zo vre-se-lijk collegiaal en lief. Altijd maak jij tijd voor mijn belangrijke en onbelangrijke vragen! Karin, ik vind het *kei spijtig* dat je ons team hebt verlaten, maar ik hoop tot in de eeuwigheid onze Wordfeud-spelletjes voort te zetten. Maria ("nee, Janneke, je doet een 'on-der-zoek', geen 'onderzoekje'! Own it."), Ruud (my brother from another Brabantse mother), Per (buum!), Monique en Nadine (mijn statistiek-hulplijnen), Diana (best roomie ever), Renske (gewoon, omdat je Renske bent), Lauraine (voor je enthousiaste ontvangst iedere dag), Marjolijn (zelfs al mis je de promotie!), Leonoor (voor onze gesprekken over mensen met 'een hoek eraf'), Jan (overbuum!), Tess (voor alle promotie-tips!), oud-CIW'ers Martijn Balsters (aka Dr. Blaster!), Mariek (gaan we een keer samen naar een 90's-party?) en Carel (het verhaal met het spiegeltje zal ik, helaas, nooit vergeten).

Ook mijn ULT-collega's verdienen een groot woord van dank. Sander (voor al het vertrouwen dat je me gaf!), Marije (alweer!), Henri, Rian, Maarten, Marloes, Vera, Vicky, Jan, Tjeerd, Joop, Arnoud-Jan, Erik, Ipek, Cathy en Minne. Het is een voorrecht en oh zo leuk om samen te werken met mensen die allemaal een enorm hart voor het onderwijs hebben. En ook dank aan alle (oud-)ULT-studenten Nederlands: het is voor mij iedere keer weer bijzonder om deel te mogen uitmaken van jullie ontwikkeling. Prachtig!

Petra, jij verdient een eigen alinea. Ik heb het je al vaker gezegd, maar nu zwart op wit: ik heb mijn hele carrière in de basis aan jou te danken. Toen ik, nog niet afgestudeerd, solliciteerde als docent NT2 bij het Talencentrum, zag je blijkbaar iets in mij en je overtuigde de rest. Jij was daarna ook degene die mij, toen ik bij de TU/e werkte, belde om voor de zekerheid toch maar even te laten weten dat er bij Geesteswetenschappen een functie vrijkwam voor een docent Academisch Nederlands. En je ziet wat daarvan is gekomen! Wacht, Petra, ik ga in jouw alinea toch ook nog even Hans van Driel noemen. Hans, jij hebt namelijk ooit precies

op het juiste moment iets belangrijks tegen mij gezegd. Jij weet misschien niet meer wat, ik wel. Ik vertel het je graag nog eens!

Fons, Anne en Kiek: de kansen die ik bij TSHD heb gekregen, zijn door jullie mogelijk gemaakt. Jullie hebben mij de ruimte gegeven om dit traject aan te gaan. En op dat punt was ik dan weer niet gekomen als Anja Klomps, destijds directeur HR, het niet had aangedurfd om mij los te laten op de TU/e: een fantastische omgeving waar ik in korte tijd enorm veel geleerd heb: op didactisch en onderwijskundig vlak maar ook om stevig in mijn schoenen te staan.

Julia, Vera, Eefje en Stefan: met het onderzoek van jullie masterscripties hebben jullie een waardevolle bijdrage geleverd aan dit proefschrift. Mijn dank daarvoor!

Naast mijn werk waren er door de jaren heen (of is het doorheen de jaren, Renske?) heel veel mensen die mijn leven op een of andere manier leuker/fijner/beter hebben gemaakt: Lotte, Roy, Arthur, Erika, Rob, Derrick, Stan, Eva, Simon, Joep, Ewen, Renske, Truyken, Anne-Marie (ook de best roomie ever), de familie van Eijk (hoi oma Liza!), de familie van der Loo, de familie van Overbeeke, mijn surrogaat-familie Leenders, de Jonosh-clan, etc. Dank jullie wel!

Kubvrouwen! Anneloes, Elise, Eugenie, Janine, Marieke en Mariken (en vooruit, een eervolle vermelding voor Hikmet): al meer dan een kwart eeuw aan uitgaan (l'ombolico del mondo!), etentjes, sinterkerstpaas-vieringen en weekendjes hebben we achter de rug, en nog altijd zijn we niet uitgepraat. Ik kijk uit naar ons volgende weekendje weg, waar we zeker weten weer als een geoliede machine de grote en kleine dingen des levens laten passeren terwijl het buiten regent. #dansendebanaan

Tinne, Tine (De VacciVodkaTinies) en Anastasia: zonder jullie was mijn integratie in België volledig mislukt. Onze vrijdagmiddagen hebben ervoor gezorgd dat ik mentaal overeind ben gebleven: we hebben uren gepraat ('gepalaaaaaaverd'), gehuild, gelachen, Verdronken Vlinder meegelald..eh…gezongen en op tafels gedanst. Met en zonder onze prachtige bende. Jullie zijn alle drie zo heerlijk jezelf. Team Rosé for life!

Natasja en Eugenie: door Corona werd de afstand Antwerpen-Bomenbuurt ineens een stuk groter. Gelukkig kan onze vriendschap na respectievelijk 39 en 26 jaar wel wat hebben. Ik kan hier geen gevatte of hilarische anekdote noemen: er zijn er te veel, waarvan er ook nog een aanzienlijk deel niet geschikt is om te delen (broccoli!). Misschien daarom eens tijd om het serieus aan te pakken: Jullie hebben me op mijn slechtst en op mijn best gezien, maar voor jullie was ik altijd goed zoals ik was. Ik geloof dat jullie al die tijd misschien wel meer vertrouwen in mij hadden dan ik in mezelf. Alleen daarom al zou iedereen een Tasja en een Eus moeten hebben. Al zijn we met z'n drieën wel bijzonder slecht in het plannen van weekendjes weg. Dus bij dezen: Barcelona, in maart? Tilburg is ook goed. April kan ook. Kijk, daar gaan we weer.

Eefje en Marloes, mijn kleine en mijn grote zus! Ik had me geen betere zussen of paranimfen (ik zal de bijnaam maar even weglaten hier) kunnen wensen. We zijn zo verschillend en toch ook weer helemaal hetzelfde. Met jullie aan mijn zijde komt het vast goed. Bedankt om er al-tijd te zijn. JeZus.

Pap en mam, er is veel gebeurd de afgelopen periode. Ik ben er ontzettend trots op dat we dit moment samen kunnen vieren. Bedankt voor de veilige basis die jullie mij (en ons Eef en Marloes) hebben gegeven. Nu is het onze beurt om er voor jullie te zijn. Het gaat ons wel lukken samen.

Ab, Lukas, Stella en Pippa: mijn dagelijkse herinnering aan wat echt belangrijk is in het leven. Ik kan nog altijd niet goed geloven dat ik het geluk heb gehad dat jullie in mijn leven zijn gekomen.

Janneke

List of Publications

Journal Articles

Van der Loo, J., Krahmer, E., & Van Amelsvoort, M. (2018). Learning How to Write an Academic Text: The Effect of Instructional Method and Writing Preference on Academic Writing Performance. *Journal of Writing Research*, 9(3), 233-258. https://doi.org/10.17239/jowr-2018.09.03.01

Van der Loo, J., Krahmer, E., & Van Amelsvoort, M. (2019). Reflection in Learning to Write an Academic Text. How does reflection affect observational learning and learning by doing in a research synthesis task? *Frontiers in Education, 4*(19). https://doi.org/10.3389/feduc.2019.00019

Van der Loo, J., Krahmer, E., & Van Amelsvoort, M. (2020). Learning How to Throw Darts. Effects of Modeling Type and Reflection on Novices' Dart-Throwing Skills. *Journal of Motor Behavio*r, 53(1), 105-116. https://doi-org.tilburguniversity.idm.oclc.org/10.1080/00222895.2020.1732861

Working Paper

Van der Loo, J., Krahmer, E., & Van Amelsvoort, M. Learning How to Combine Sentences. Effects of Modeling Type and Reflection on Writing Revision Proficiency. *Manuscript submitted for publication.*

Conference Proceedings

Van der Loo, J., Krahmer E., & Van Amelsvoort (2016). Learning How To Write an Academic Text: the Effect of Instructional Method and Reflection on Text Quality.
Proceedings of the 38th International Conference on Cognition and Exploratory Learning in Digital Age. Mannheim, Germany.

Van der Loo, J., Frissen, E., & Krahmer, E. (2016). Learning How To Throw Darts: The Effect Of Modeling Type And Reflection On Dart-Throwing Skills. *Proceedings of the 38th Annual Meeting of the Cognitive Science Society.* Philadelphia, United States.

Conijn, R., **van der Loo, J**., & Van Zaanen, M. (2018, March). What's (not) in a keystroke? Automatic discovery of students' writing processes using keystroke logging. *Proceedings of the 8th International Conference on Learning Analytics and Knowledge.* Sydney, Australia.

Bouwer, R., & **van der Loo, J.** (2021). Grip op je schrijfproces met De Schrijfmachine: Zelfregulerend academisch schrijven. *Proceedings of the 34th HSN Conferentie Onderwijs Nederlands.* Online, The Netherlands.

Conference Presentations

Van der Loo, J. (2017, August 29 – September 2). *Learning to Write a Synthesis Text: The Effect of Instructional Method and Reflection.* 17th Biennial EARLI Conference, Tampere, Finland.

Van der Loo, J. (2018). Effect of Modeling Type and Reflection in a Sentence Combining Task. (2018, August 29 – 31). SIG Writing Conference, Antwerp, Belgium.

Bouwer, R., & **van der Loo, J.** (July 7-9, 2021). *Grip op je schrijfproces met De Schrijfmachine: Zelfregulerend academisch schrijven*. Onderwijs Research Dagen, online, the Netherlands.

TiCC Ph.D. Series

1. Pashiera Barkhuysen. Audiovisual Prosody in Interaction. Promotores: M.G.J. Swerts, E.J. Krahmer. Tilburg, 3 October 2008.
2. Ben Torben-Nielsen. Dendritic Morphology: Function Shapes Structure. Promotores: H.J. van den Herik, E.O. Postma. Co-promotor: K.P. Tuyls. Tilburg, 3 December 2008.
3. Hans Stol. A Framework for Evidence-based Policy Making Using IT. Promotor: H.J. van den Herik. Tilburg, 21 January 2009.
4. Jeroen Geertzen. Dialogue Act Recognition and Prediction. Promotor: H. Bunt. Co-promotor: J.M.B. Terken. Tilburg, 11 February 2009.
5. Sander Canisius. Structured Prediction for Natural Language Processing. Promotores: A.P.J. van den Bosch, W. Daelemans. Tilburg, 13 February 2009.
6. Fritz Reul. New Architectures in Computer Chess. Promotor: H.J. van den Herik. Co-promotor: J.W.H.M. Uiterwijk. Tilburg, 17 June 2009.
7. Laurens van der Maaten. Feature Extraction from Visual Data. Promotores: E.O. Postma, H.J. van den Herik. Co-promotor: A.G. Lange. Tilburg, 23 June 2009 (cum laude).
8. Stephan Raaijmakers. Multinomial Language Learning. Promotores: W. Daelemans, A.P.J. van den Bosch. Tilburg, 1 December 2009.
9. Igor Berezhnoy. Digital Analysis of Paintings. Promotores: E.O. Postma, H.J. van den Herik. Tilburg, 7 December 2009.
10. Toine Bogers. Recommender Systems for Social Bookmarking. Promotor: A.P.J. van den Bosch. Tilburg, 8 December 2009.
11. Sander Bakkes. Rapid Adaptation of Video Game AI. Promotor: H.J. van den Herik. Co-promotor: P. Spronck. Tilburg, 3 March 2010.
12. Maria Mos. Complex Lexical Items. Promotor: A.P.J. van den Bosch. Co-promotores: A. Vermeer, A. Backus. Tilburg, 12 May 2010 (in collaboration with the Department of Language and Culture Studies).
13. Marieke van Erp. Accessing Natural History. Discoveries in data cleaning, structuring, and retrieval. Promotor: A.P.J. van den Bosch. Co-promotor: P.K. Lendvai. Tilburg, 30 June 2010.
14. Edwin Commandeur. Implicit Causality and Implicit Consequentiality in Language Comprehension. Promotores: L.G.M. Noordman, W. Vonk. Co-promotor: R. Cozijn. Tilburg, 30 June 2010.
15. Bart Bogaert. Cloud Content Contention. Promotores: H.J. van den Herik, E.O. Postma. Tilburg, 30 March 2011.
16. Xiaoyu Mao. Airport under Control. Promotores: H.J. van den Herik, E.O. Postma. Co-promotores: N. Roos, A. Salden. Tilburg, 25 May 2011.
17. Olga Petukhova. Multidimensional Dialogue Modelling. Promotor: H. Bunt. Tilburg, 1 September 2011.

18. Lisette Mol. Language in the Hands. Promotores: E.J. Krahmer, A.A. Maes, M.G.J. Swerts. Tilburg, 7 November 2011 (cum laude).
19. Herman Stehouwer. Statistical Language Models for Alternative Sequence Selection. Promotores: A.P.J. van den Bosch, H.J. van den Herik. Co-promotor: M.M. van Zaanen. Tilburg, 7 December 2011.
20. Terry Kakeeto-Aelen. Relationship Marketing for SMEs in Uganda. Promotores: J. Chr. van Dalen, H.J. van den Herik. Co-promotor: B.A. Van de Walle. Tilburg, 1 February 2012.
21. Suleman Shahid. Fun & Face: Exploring Non-Verbal Expressions of Emotion during Playful Interactions. Promotores: E.J. Krahmer, M.G.J. Swerts. Tilburg, 25 May 2012.
22. Thijs Vis. Intelligence, Politie en Veiligheidsdienst: Verenigbare Grootheden? Promotores: T.A. de Roos, H.J. van den Herik, A.C.M. Spapens. Tilburg, 6 June 2012 (in collaboration with the Tilburg School of Law).
23. Nancy Pascall. Engendering Technology Empowering Women. Promotores: H.J. van den Herik, M. Diocaretz. Tilburg, 19 November 2012.
24. Agus Gunawan. Information Access for SMEs in Indonesia. Promotor: H.J. van den Herik. Co-promotores: M. Wahdan, B.A. Van de Walle. Tilburg, 19 December 2012.
25. Giel van Lankveld. Quantifying Individual Player Differences. Promotores: H.J. van den Herik, A.R. Arntz. Co-promotor: P. Spronck. Tilburg, 27 February 2013.
26. Sander Wubben. Text-to-text Generation Using Monolingual Machine Translation. Promotores: E.J. Krahmer, A.P.J. van den Bosch, H. Bunt. Tilburg, 5 June 2013.
27. Jeroen Janssens. Outlier Selection and One-Class Classification. Promotores: E.O. Postma, H.J. van den Herik. Tilburg, 11 June 2013.
28. Martijn Balsters. Expression and Perception of Emotions: The Case of Depression, Sadness and Fear. Promotores: E.J. Krahmer, M.G.J. Swerts, A.J.J.M. Vingerhoets. Tilburg, 25 June 2013.
29. Lisanne van Weelden. Metaphor in Good Shape. Promotor: A.A. Maes. Co-promotor: J. Schilperoord. Tilburg, 28 June 2013.
30. Ruud Koolen. Need I say More? On Overspecification in Definite Reference. Promotores: E.J. Krahmer, M.G.J. Swerts. Tilburg, 20 September 2013.
31. J. Douglas Mastin. Exploring Infant Engagement. Language Socialization and Vocabulary. Development: A Study of Rural and Urban Communities in Mozambique. Promotor: A.A. Maes. Co-promotor: P.A. Vogt. Tilburg, 11 October 2013.
32. Philip C. Jackson. Jr. Toward Human-Level Artificial Intelligence – Representation and Computation of Meaning in Natural Language. Promotores: H.C. Bunt, W.P.M. Daelemans. Tilburg, 22 April 2014.
33. Jorrig Vogels. Referential Choices in Language Production: The Role of Accessibility. Promotores: A.A. Maes, E.J. Krahmer. Tilburg, 23 April 2014 (cum laude).
34. Peter de Kock. Anticipating Criminal Behaviour. Promotores: H.J. van den Herik, J.C. Scholtes. Co-promotor: P. Spronck. Tilburg, 10 September 2014.
35. Constantijn Kaland. Prosodic Marking of Semantic Contrasts: Do Speakers Adapt to Addressees? Promotores: M.G.J. Swerts, E.J. Krahmer. Tilburg, 1 October 2014.
36. Jasmina Marić. Web Communities, Immigration and Social Capital. Promotor: H.J. van den Herik. Co-promotores: R. Cozijn, M. Spotti. Tilburg, 18 November 2014.

37. Pauline Meesters. Intelligent Blauw. Promotores: H.J. van den Herik, T.A. de Roos. Tilburg, 1 December 2014.
38. Mandy Visser. Better Use Your Head. How People Learn to Signal Emotions in Social Contexts. Promotores: M.G.J. Swerts, E.J. Krahmer. Tilburg, 10 June 2015.
39. Sterling Hutchinson. How Symbolic and Embodied Representations Work in Concert. Promotores: M.M. Louwerse, E.O. Postma. Tilburg, 30 June 2015.
40. Marieke Hoetjes. Talking hands. Reference in Speech, Gesture and Sign. Promotores: E.J. Krahmer, M.G.J. Swerts. Tilburg, 7 October 2015.
41. Elisabeth Lubinga. Stop HIV. Start Talking? The Effects of Rhetorical Figures in Health Messages on Conversations among South African Adolescents. Promotores: A.A. Maes, C.J.M. Jansen. Tilburg, 16 October 2015.
42. Janet Bagorogoza. Knowledge Management and High Performance. The Uganda Financial Institutions Models for HPO. Promotores: H.J. van den Herik, B.A. van de Walle. Tilburg, 24 November 2015.
43. Hans Westerbeek. Visual realism: Exploring Effects on Memory, Language Production, Comprehension, and Preference. Promotores: A.A. Maes, M.G.J. Swerts. Co-promotor: M.A.A. van Amelsvoort. Tilburg, 10 February 2016.
44. Matje van de Camp. A link to the Past: Constructing Historical Social Networks from Unstructured Data. Promotores: A.P.J. van den Bosch, E.O. Postma. Tilburg, 2 March 2016.
45. Annemarie Quispel. Data for all: How Professionals and Non-Professionals in Design Use and Evaluate Information Visualizations. Promotor: A.A. Maes. Co-promotor: J. Schilperoord. Tilburg, 15 June 2016.
46. Rick Tillman. Language Matters: The Influence of Language and Language Use on Cognition. Promotores: M.M. Louwerse, E.O. Postma. Tilburg, 30 June 2016.
47. Ruud Mattheij. The Eyes Have It. Promotores: E.O. Postma, H. J. Van den Herik, and P.H.M. Spronck. Tilburg, 5 October 2016.
48. Marten Pijl. Tracking of Human Motion over Time. Promotores: E. H. L. Aarts, M. M. Louwerse. Co-promotor: J. H. M. Korst. Tilburg, 14 December 2016.
49. Yevgen Matusevych. Learning Constructions from Bilingual Exposure: Computational Studies of Argument Structure Acquisition. Promotor: A.M. Backus. Co-promotor: A. Alishahi. Tilburg, 19 December 2016.
50. Karin van Nispen. What Can People with Aphasia Communicate with their Hands? A Study of Representation Techniques in Pantomime and Co-Speech Gesture. Promotor: E.J. Krahmer. Co-promotor: M. van de Sandt-Koenderman. Tilburg, 19 December 2016.
51. Adriana Baltaretu. Speaking of Landmarks. How Visual Information Influences Reference in Spatial Domains. Promotores: A.A. Maes and E.J. Krahmer. Tilburg, 22 December 2016.
52. Mohamed Abbadi. Casanova 2, a Domain Specific Language for General Game Development. Promotores: A.A. Maes, P.H.M. Spronck, A. Cortesi. Co-promotor: G. Maggiore. Tilburg, 10 March 2017.

53. Shoshannah Tekofsky. You Are Who You Play You Are. Modelling Player Traits from Video Game Behavior. Promotores: E.O. Postma, P.H.M. Spronck. Tilburg, 19 June 2017.
54. Adel Alhuraibi. From IT-BusinessStrategic Alignment to Performance: A Moderated Mediation Model of Social Innovation, and Enterprise Governance of IT. Promotores: H.J. van den Herik, B.A. van de Walle. Co-promotor: S. Ankolekar. Tilburg, 26 September 2017.
55. Wilma Latuny. The Power of Facial Expressions. Promotores: E.O. Postma, H.J. van den Herik. Tilburg, 29 September 2017.
56. Sylvia Huwaë. Different Cultures, Different Selves? Suppression of Emotions and Reactions to Transgressions across Cultures. Promotores: E.J. Krahmer, J. Schaafsma. Tilburg, 11 October 2017.
57. Mariana Serras Pereira. A Multimodal Approach to Children's Deceptive Behavior. Promotor: M. Swerts. Co-promotor: S. Shahid. Tilburg, 10 January 2018.
58. Emmelyn Croes. Meeting Face-to-Face Online: The Effects of Video-Mediated Communication on Relationship Formation. Promotores: E.J. Krahmer, M. Antheunis. Co-promotor: A.P. Schouten. Tilburg, 28 March 2018.
59. Lieke van Maastricht. Second language prosody: Intonation and rhythm in production and perception. Promotores: E.J. Krahmer, M.G.J. Swerts. Tilburg, 9 May 2018.
60. Nanne van Noord. Learning visual representations of style. Promotores: E.O. Postma, M. Louwerse. Tilburg, 16 May 2018.
61. Ingrid Masson Carro. Handmade: On the cognitive origins of gestural representations. Promotor: E.J. Krahmer. Co-promotor: M.B. Goudbeek. Tilburg, 25 June 2018.
62. Bart Joosten. Detecting social signals with spatiotemporal Gabor filters. Promotores: E.J. Krahmer, E.O. Postma. Tilburg, 29 June 2018.
63. Yan Gu. Chinese hands of time: The effects of language and culture on temporal gestures and spatio-temporal reasoning. Promotor: M.G.J. Swerts. Co-promotores: M.W. Hoetjes, R. Cozijn. Tilburg, 5 June 2018.
64. Thiago Castro Ferreira. Advances in natural language generation: Generating varied outputs from semantic inputs. Promotor: E.J. Krahmer. Co-promotor: S. Wubben. Tilburg, 19 September 2018.
65. Yu Gu. Automatic emotion recognition from Mandarin speech. Promotores: E.O. Postma, H.J. van den Herik, H.X. Lin. Tilburg, 28 November 2018.
66. Francesco Di Giacomo. Metacasanova: A high-performance meta-compiler for domain-specific languages. Promotores: P.H.M Spronck, A. Cortesi, E.O. Postma. Tilburg, 19 November 2018.
67. Ákos Kádár. Learning visually grounded and multilingual representations. Promotores: E.O. Postma. Co-promotor: A. Alishahi, G.A. Chrupala. Tilburg, 13 November 2019.
68. Phoebe Mui. The many faces of smiling: Social and cultural factors in the display and perception of smiles. Promotor: M.G.J. Swerts. Co-promotor: M.B. Goudbeek. Tilburg, 18 December 2019.
69. Véronique Verhagen. Illuminating variation: Individual differences in entrenchment of multi-word units. Promotor: A.M. Backus. Co-promotores: M.B.J. Mos, J. Schilperoord. Tilburg, 10 January 2020 (cum laude).

70. Debby Damen. Taking perspective in communication: Exploring what it takes to change perspectives. Promotor: E.J. Krahmer. Co-promotores: M.A.A. Van Amelsvoort, P.J. Van der Wijst. Tilburg, 4 November 2020.
71. Alain Hong. Women in the Lead: Gender, Leadership Emergence, and Negotiation Behavior from a Social Role Perspective. Promotor: J. Schaafsma. Co-promotor: P.J. van der Wijst. Tilburg, 3 June 2020.
72. Chrissy Cook. Everything You Never Wanted to Know about Trolls: An Interdisciplinary Exploration of the Who's, What's and Why's of Trolling in Online Games. Promotores: J. Schaafsma, M.L. Antheunis. Tilburg, 22 January 2021.
73. Nadine Braun. Affective Words and the Company They Keep: Investigating the interplay of emotion and language. Promotor: E.J. Krahmer. Co-promotor: M.B. Goudbeek. Tilburg, 29 March 2021.
74. Yueqiao Han. Chinese Tones: Can You Listen with Your Eyes? The Influence of Visual Information on Auditory Perception of Chinese Tones. Promotor: M.G.J. Swerts. Co-promotores: M.B.J. Mos, M.B. Goudbeek. Tilburg, 18 June 2021.
75. Tess van der Zanden. Language Use and Impression Formation: The Effects of Linguistic Cues in Online Dating Profiles. Promotor: E.J. Krahmer. Co-promotores: M.B.J. Mos, A.P. Schouten. Tilburg, 22 October 2021.
76. Janneke van der Loo. Mastering the Art of Academic Writing: Comparing the Effectiveness of Observational Learning and Learning by Doing. Promotor: E.J. Krahmer. Co-promotor: M.A.A. van Amelsvoort. Tilburg, 1 December 2021.