Professional development of teachers in the implementation of a strategy-focused writing intervention program for elementary students

Professional Development of Teachers in the Implementation of a Strategy-Focused Writing Intervention Program for Elementary Students

Monica Koster, Renske Bouwer, and Huub van den Bergh

Utrecht University

Submitted: January 18, 2016

Revised & resubmitted: June 10, 2016

Abstract

In this study we examined the effectiveness of *Tekster* [Texter], a comprehensive program for writing for the upper elementary grades, combining strategy instruction, text structure instruction, and the teaching of self-regulation skills with observational learning, explicit instruction, and (guided) practice to address both the focus of instruction (what is taught) and the mode of instruction (how it is taught). Further, we investigated the added value of a professional development program for teachers on the effectiveness and implementation of the intervention in the classroom, by adopting a teachers-training-teachers approach. One group of teachers (N=31) was trained by experts, and subsequently trained their colleagues (N=37). Quasi-experimental results showed that students' writing performance improved after the intervention (ES = 0.55), while generalizing over tasks, students, and teachers. Further, teachers became more positive and felt more efficacious about teaching writing after the intervention. There were no differences between trainers and trainees, which provides evidence for the spillover effect of professional development. To get more insight in how teachers implemented the intervention in their classroom and in the social validity of the intervention and the teachers-training-teachers approach, we triangulated post-intervention questionnaires with classroom observations and interviews. This mixed methods approach revealed that both trainers and trainees were highly satisfied with the program and easily adapted their focus of instruction. However, for adjusting the mode of instruction more teacher support seems to be needed.

Keywords: teaching writing, professional development, elementary students, strategy instruction, mixed methods

1. Introduction

In many countries concerns are raised about the level of writing proficiency of elementary students (cf. Department for Education, 2012; Salahu-Din, Persky, & Miller, 2008). In the Netherlands has also been established that students' writing performance at the end of elementary school does not meet the standards set by the Ministery of Education (Henkens, 2010). As a target goal for the end of elementary school the Ministry proposes that "students are able to write coherent texts, with a simple linear structure on various familiar topics; the text includes an introduction, body, and ending" (Expert Group Learning Trajectories, 2009, p.15). However, at the end of elementary school the majority of Dutch students is not capable of composing a text that successfully conveys a message to a reader (Kuhlemeier, Til, Hemker, De Klijn, & Feenstra, 2013). Why is writing so hard for elementary students? The major problem developing writers face during writing is cognitive overload. Writing is a complex cognitive process, during which several resource-demanding cognitive activities have to be performed simultaneously, such as activating prior knowledge, generating content, planning, formulating, and revising, whilst taking into account the communicative goal of the text and the intended audience (Fayol, 1999). Additionally, the amount of attention required for foundational skills (e.g., handwriting, spelling, and sentence construction) needs to be considered. This is particularly relevant with developing writers, as they often lack automaticity in these areas (McCutchen, 2011). Due to this limited automaticity, the learner has less attentional capacity for the higher level processes in writing, such as planning, formulating, and revising, which has detrimental effects on text quality (Berninger, Yates, Cartwright, Rutberg, Remy, & Abbott, 1992; McCutchen, 1996). An additional source of cognitive overload is the fact that, in the way writing education is often organized, learning-to-write and task execution are inextricably linked. For novice

writers text production is already so cognitively demanding, that there is hardly any attentional capacity left for learning (Rijlaarsdam & Couzijn, 2000). Thus, writing instruction should aim to improve students' writing performance by teaching them skills and knowledge to manage the cognitive activities during writing. To achieve this, writing instruction needs to address the focus of instruction (what is taught) as well as the mode of instruction (how it is taught).

1.1. Writing instruction: The present situation

The Dutch Inspectorate for the Education reported that in the average classroom attention and time devoted to writing are limited, and that the majority of teachers do not succeed in effectively teaching writing (Henkens, 2010). There are two reasons for these shortcomings in writing education: (1) a lack of suitable teaching materials, and (2) teachers lack the necessary skills and knowledge for effectively teaching writing (Pullens, 2012; Van der Leeuw, 2006). Teachers often do not explain how students can approach a writing task, discuss texts, provide feedback, nor do they promote rereading and revising activities (Henkens, 2010). Although the language teaching materials pay attention to process-directed writing education, they do not offer teachers enough support to adequately assist their students during the writing process. Support for teachers is essential, as during their preservice and in-service professional development they are not sufficiently prepared to teach writing (Pullens, 2012; Van der Leeuw, 2006). Time devoted to the didactics of writing is limited, and student-teachers are expected to acquire the required skills and knowledge independently through learning-by-doing. As part of their training prospective teachers have to write a lot, but due to limited time and resources, they hardly receive any feedback on their writing (Van der Leeuw, Pauw, Smits, & Van de Ven, 2010). Thus, not only teaching materials need to be improved, but

also the skills and knowledge of teachers need to be extended to optimize the focus and mode of writing instruction in elementary school. Already a lot of research has been done on both these aspects, identifying several effective instructional practices. These will be discussed below.

1.2. Optimizing the focus of instruction

Concerning the focus of instruction, several meta-analyses have identified various effective instructional practices to enhance students' writing performance, such as strategy instruction, teaching students self-regulation skills for writing, and text structure instruction, (Graham, 2006; Graham, McKeown, Kiuhara, & Harris, 2012; Koster, Tribushinina, De Jong, & Van den Bergh, 2015). Teaching students to adopt strategies before, during and after writing is an effective way to reduce cognitive overload during writing as this limits the number of cognitive processes that are active at the same time (Kellogg, 1988, 2008). For example, when students are taught to plan during the prewriting phase, they can focus on non-planning processes during writing. Studies involving explicit strategy instruction invariably yield large effect sizes, ranging from 0.82 to 1.15 (Graham, 2006; Graham, McKeown, Kiuhara, & Harris, 2012; Graham & Perin, 2007; Hillocks, 1984; Koster, Tribushinina, De Jong, & Van den Bergh, 2015).

The combination of strategy instruction with teaching self-regulatory skills yields an even higher effect size, ES = 1.17 (Graham et al., 2012). Essential self-regulatory skills in writing are setting goals for writing, and subsequently monitoring the progress towards these goals (Flower & Hayes, 1981). The most prominent and well-researched approach combining strategy instruction and the teaching of self-regulation skills is the Self-Regulated Strategy Development (SRSD) (Harris, Graham, Mason, & Saddler, 2002). In SRSD students are taught strategies for planning, writing, revising and editing,

and they are supported in the development of the self-regulation procedures needed to monitor and manage their writing. This instructional approach has been implemented in small groups and whole classrooms with students of different age groups and abilities, and has invariably proven to be very effective in improving students' writing performance (Harris et al., 2002).

To be able to set effective goals for writing, students need to know what communicative goals should be set for which type of text and how you write a text meeting these goals. For this, students need to have knowledge about text structures and criteria for a good text. The effect of explicit text structure instruction, in which the elements and organization of text types are specifically taught, has been extensively examined in the elementary grades, in different genres: narrative (Fitzgerald & Teasley, 1986; Gordon & Braun, 1986), persuasive (Crowhurst, 1990, 1991; Scardamalia & Paris, 1985), and informative (Bean & Steenwyk, 1984; Raphael & Kirschner, 1985). Meta-analyses (Graham et al., 2012; Koster et al., 2015) show that the overall effect of text structure instruction was positive (ESs 0.59 and 0.76 respectively).

1.3. Optimizing the mode of instruction

Writing instruction must be optimized to address the double challenge problem of learning-to-write and task execution. An effective approach to separate these two components and provide students with the opportunity to fully direct their attention to learning- to-write is observational learning (Zimmerman & Risemberg, 1997). By observing a model performing (part of) a writing task while explaining, demonstrating, and verbalizing his thoughts, students gain insight into the writing process. This prepares them for the writing task and supports them during their writing process (Rijlaarsdam & Couzijn, 2000). Various studies have demonstrated the effectiveness of teacher modeling

as an instructional mode to teaching writing strategies (cf. Graham, Harris, & Mason, 2005; Fidalgo, Torrance, Rijlaarsdam, Van den Bergh, & Lourdes Álvarez, 2015). Peers can also be used as models (cf. Braaksma, Rijlaarsdam, Van den Bergh, & Van Hout-Wolters, 2010). Besides positive effects on students' writing performances and writing processes (Braaksma, 2002; Braaksma et al., 2010), peer modeling also has positive effects on self-efficacy and motivation, especially in weaker students (Schunk, 1987).

Observational learning can also be applied by confronting students with reader reactions to provide them feedback on the communicative effectiveness of the written product (cf. Couzijn & Rijlaarsdam, 2004; Holliway & McCutchen, 2004). Beginning writers often are unaware of the communicative deficiencies in their writing. Observing genuine readers and discussing readers' experiences provide students with valuable information on the readers' needs and whether they succeeded in fulfilling these needs (Couzijn & Rijlaarsdam, 2004; Schriver, 1992). Several researchers (Couzijn, 1995; Couzijn & Rijlaarsdam, 2004; Holliway & McCutchen, 2004; Rijlaarsdam, Couzijn, Janssen, Braaksma, & Kieft, 2006) have shown that students' writing improved when they experience the effect their text has on a reader.

Although observational learning is effective in improving students' writing, there is still a gap to be bridged: from observing to independent practice. The teacher can facilitate student's progress through scaffolding with a gradual release of responsibility. In scaffolding the teacher controls the elements of the task that are initially beyond the student's capacity, thus permitting the student to concentrate upon the elements that are within his range of competence (Wood, Brunner, & Ross, 1976). The amount of teacher assistance can gradually be decreased as the learner progresses, and through guided practice and, finally, independent performance the cognitive load shifts from teacher to student (Pearson & Gallagher, 1983; Wood et al., 1976). Intervention programs that use

gradual release of responsibility and scaffolding have been successful in improving students' writing performance (cf. Graham, MacArthur, & Schwartz, 1995; Graham et al., 2005).

1.4. Bringing writing research into the classroom

1.4.1. Teacher involvement in research

Although the last decades of writing intervention research have provided guidance for the improvement of the teaching of writing, the actual implementation of evidence-based instructional practices is arduous, due to a substantial gap between research and classroom practice (Broekkamp & Van Hout-Wolters, 2007). To bridge this gap and effectively improve classroom practice, it is essential to involve teachers in intervention studies in a meaningful way (Borko, 2004). In the intervention studies (N = 32) that were analyzed in the meta-analysis of Koster and colleagues (2015), regular classroom teachers were not involved in the research in nearly half of the studies; the intervention was delivered either by the researchers themselves, or by trained research assistants. The results of these studies show positive effects on students' performance, but one can hardly expect any improvement after the intervention, as there is no encouragement or support for teachers to change the way they teach writing. This is also the case for 12% of the studies in which the teacher delivered the intervention with materials supplied by the researcher, without any additional training. Once the intervention has ended, teachers will return to working with their regular materials, so here also no longer lasting intervention effects are to be expected.

In the remaining 40% of the sample, teachers deliver the intervention themselves, either after having received a (short) training (21%) or a more extensive form of professional development (12%). Lastly, in 6% of the studies the teacher is part of the

research team. Only under these circumstances, a change of teachers' instructional practices may be expected. It is worth mentioning that in this sample of studies the average effect size does not differ significantly between studies in which teachers were involved (ES = 0.79, *SD* = 0.86) versus studies in which they were not (ES = 0.99, *SD* = 0.74, $t(31) = .70$, $p = .49$). Thus, the inclusion of teachers in research does not seem to lead to a significant decrease in effect sizes. This pleads for inclusion of teachers in research.

1.4.2. Professional development

To meaningfully involve teachers in intervention research and improve classroom practice, teachers should be provided with the prerequisite tools to successfully implement the intervention. Therefore it is important that intervention studies include professional development activities for teachers. An effective way to organize professional learning for teachers is the practice-based professional development approach (Ball & Forzani, 2009). Practice-based professional development focuses on developing teachers' understanding and skills to effectively implement an educational practice, instead of focusing primarily on increasing teachers' knowledge about a practice (Ball & Forzani, 2009). Important features that have proven to be effective for teachers' professional development are (a) consistency with existing knowledge and beliefs, (b) focus on content and how students learn that content, (c) alignment with state standards, (d) opportunities for teachers to engage in active learning and (e) collaboration between teachers (Desimone, 2009; Harris, Lane, Graham, Driscoll, Sandmel, Brindle, & Schatschneider, 2012). Collaboration can be promoted by collective participation of teachers in schools, and by providing time and space for sharing, observing expert teachers in practice, being observed and receiving feedback, for instance in professional

learning communities (Borko, 2004; Guskey, 1994; Harris et al., 2012). It has been established that professional learning communities lead to increased involvement, ownership, innovation, and leadership among teachers (Borko, 2004). Finally, professional development activities or programs should be sufficient in duration: this concerns the actual number of hours spent as well as the time span over which the trajectory is spread (Desimone, 2009).

1.4.3. Teachers' self-efficacy for teaching writing

Improved teacher practices have a positive influence on student achievement (Desimone, 2009). As it is important that teachers are confident that they can affect students' learning outcomes, professional development should not only address the skills and knowledge that are required for successful and effective writing instruction, but also the beliefs about and attitudes toward writing instruction (Graham, Harris, Fink, & MacArthur, 2001). The beliefs that teachers hold about their ability to teach writing affect how they use the skills and knowledge that they have about teaching writing during their writing instruction (Graham et al., 2001; Pajares, 1992; Rietdijk, Van Weijen, Janssen, Van den Bergh, & Rijlaarsdam, 2015), which influences the overall quality of the instruction (Tschannen-Moran, Woolfolk Hoy, & Hoy, 1998). De Smedt, Van Keer, and Merchie (2016) found that teachers' efficacy for writing was positively related with students' writing performance. Thus, a higher feeling of self-efficacy of teachers for (teaching) writing results in a higher quality writing instruction, which leads to better student performance. This suggests that teachers' self-efficacy is a key factor in the improvement of writing education and that to improve the quality of teachers' instruction it is essential to enhance their feeling of self-efficacy by training them in applying effective writing practices (De Smedt et al., 2016).

1.4.4. Social validity

Another essential aspect in bridging the gap between research and practice is the social validity of an intervention, i.e., the acceptability of and satisfaction with the intervention procedures according to the individuals who receive and implement the intervention procedures (Luiselli & Reed, 2011). In the case of an intervention aimed to improve classroom practice, teachers might be queried about the complexity of the followed procedure, time involved with the implementation of the intervention, and satisfaction with the outcome (Luiselli & Reed, 2011). This yields important information about the feasibility of the implementation of an intervention in daily classroom practice. Procedures that are perceived as too complicated, impractical or unhelpful, will likely not be adopted. Social validity is a key aspect in the long term effects of interventions: a higher social validity increases the likelihood that (aspects of) an intervention will still be applied after the intervention period has ended. Social validity can be assessed through interviews, surveys, or questionnaires. However, each of these measures separately only provides information about a single aspect of the intervention. To obtain an impression of the full potential of the usability of an intervention in daily classroom practice, the results of the various social validity measures should be combined, for instance by using a mixed methods approach (Luiselli & Reed, 2011).

1.5. Testing the effectiveness of a writing intervention program in the classroom

The need to include professional development activities for teachers in the implementation of an intervention program was illustrated by the results of a recent intervention study, in which the effectiveness of a newly developed comprehensive program for teaching writing was examined (Bouwer, Koster, & Van den Bergh, 2016a).

This program, called *Tekster* [Texter] (Bouwer et al, 2016a; Koster, Bouwer, & Van den Bergh, 2014a, 2014b, 2014c), aimed to improve the writing performance of students in the upper grades of elementary school in the Netherlands, and combined several effective instructional practices into one general overall approach for writing. In *Tekster* the main focus of instruction was to teach students a strategy for writing, based on the steps of the writing process, i.e., planning, writing, and revising. The main focus in grade 4 was on prewriting activities (generate and organize content), this shifted to post-writing activities (evaluating and revising) in grade 6. Strategy instruction was supplemented with explicit instruction in text structure and the teaching of self-regulatory skills. The predominant mode of instruction was observational learning, complemented with explicit instruction and (guided) practice with extensive scaffolding, following the gradual release of responsibility model.

Tekster was tested in a large intervention study in a natural setting, with 60 teachers and 1420 students. The intervention was delivered by the teachers themselves, after only one short introductory training session. Results showed that the program was already effective over a period of two months: students' writing performance improved significantly across all grades (ES = 0.40), whilst generalizing over tasks, and this improvement was maintained two months after the intervention (Bouwer, et al, 2016a). However, the intraclass correlation was .37, indicating that there was a large proportion of variance attributable to classes: in some classes students hardly made any progress in their writing performance, whereas in other classes students progressed a full grade level. These large differences between classes indicate that teachers need more support in implementing the program more effectively and with more fidelity. A supplementary professional development program could offer the support teachers need and might minimize differences between teachers.

However, in a large-scale intervention study it is not always a feasible option to have all teachers involved in the study participate in an extensive professional development program. It is therefore necessary to examine alternative ways in which teachers can benefit from professional development. Sun, Penuel, Frank, Gallagher, and Youngs (2013) have investigated the so-called *spillover effect*, which means that through collegial interactions teachers can learn from professional development participants. The results from this study are promising: spillover effects can be almost as large as the direct effects of professional development. In the US National Writing Project (NWP) a similar method was used to reach large numbers of teachers: a teachers-training-teachers approach (Borko, 2004; Lieberman & Friedrich, 2007). In this project, teachers attended summer institutes and subsequently provided workshops for their colleagues. These teacher-trainers reported that it took time to earn their leadership, but that they eventually succeeded in doing so by showing their commitment, by high-quality teaching, and their willingness to give advice to their colleagues (Lieberman & Friedrich, 2007). Teachers reported that NWP has brought about change in their beliefs and attitudes about teaching writing, and the majority of students' work showed improvements in organization, coherence and use of writing conventions (Borko, 2004). Thus, a teacher-training-teacher approach is a promising method to train large numbers of teachers.

1.6. Aim of the present study

In the present study, we examined the effect of the writing intervention program *Tekster,* including professional development activities for teachers, on students' writing performance, and on teachers' self-efficacy and attitudes for writing and the teaching of writing. Results from a previous intervention study showed that the program was effective, but large differences between teachers were an indication that more support in

implementing the program was required (Bouwer et al, 2016a). Therefore, in this study, we included professional development activities to offer teachers support in the implementation of *Tekster* and increase their skills and knowledge of teaching writing. To investigate whether the content of such a professional development program could be transferred between teachers, we applied a teachers-training-teachers approach in which half of the participating teachers were trained by the researchers, and these teacher-trainers subsequently trained one or more colleagues.

We investigated the effect of the intervention on student outcomes, and examined whether there were differences between students in the group of teachers who were trained by experts (teacher-trainers), and teachers who were trained by colleagues (teacher-trainees), and whether there were differences between grades. Further, we examined the effect of the intervention program and professional development activities on the self-efficacy and attitudes of teachers towards writing and teaching writing and whether there were differences between teacher-trainers, and teacher-trainees. The effect of the intervention on students' writing performance was examined by a quasi-experimental design with two groups (teacher-trainers versus teacher-trainees) and three measurement occasions. The first measurement occasion was a pretest for both groups. The first group (teacher-trainers) received the intervention during the first time interval, from the first to the second measurement occasion. The second measurement occasion served as a posttest for this group, and as a second pretest for the second group (teacher-trainees). Between the second and third measurement occasion, the second group received the intervention. Therefore, the third measurement occasion served as a posttest for the second group, as well as a delayed posttest for the second group, which provided information on the long-term effects of the intervention. To examine the effect of the intervention on the self-efficacy and attitudes of teachers, and whether this effect differed

between teacher-trainers and teacher-trainees, we administered questionnaires prior to and after the intervention.

We expected that students' writing would improve after the intervention in both groups, but not necessarily to the same extent, as research has shown that spillover effects can be *almost* as large as the direct effects of professional development (Sun et al., 2013). Regarding the differences between grades we expect that students in grade 6 will write qualitatively better texts than students in grade 4, as they have had more years of schooling and practice. We expect improvements in self-efficacy, and that all teachers become more positive about writing and writing instruction due to the professional development program, but again, not necessarily to the same extent.

As teachers were a crucial factor in the implementation of the intervention program, it was important to establish how they have implemented the program and if they implemented the program with fidelity. For this, we specifically examined how teachers implemented key components of the intervention program in their writing lessons and how trainers transferred their knowledge and skills to colleagues during the collegial training sessions. Lastly, we investigated the social validity of the intervention program and the teachers-training-teachers approach, as this provides valuable information about the feasibility of the implementation of the intervention in daily classroom practice. With regards to social validity, we examined teachers' experiences and satisfaction with the lesson material, the teacher manual, and the training sessions. To examine the implementation and the social validity of the intervention program including the professional development activities, we used an explanatory sequential mixed methods approach in which we collected and analyzed quantitative data from post-intervention questionnaires, logbooks and classroom observations, and followed up with qualitative data from focus group interviews to further explain how teachers implemented

and experienced the intervention program and to elaborate on differences between trainers and trainees. This will provide us with valuable clues on how the program was implemented in schools.

2. Method

2.1. Sample

In total, 68 teachers and 1365 students from 65 classes and 25 elementary schools participated. All teachers were qualified and experienced elementary teachers. On average, they had 11.22 years (SD = 8.50) of experience in elementary grades, with an average of 3.55 years in the grade in which they were teaching at the start of the study. The majority of teachers were female (81%) and they were from schools spread all over the country: 10 schools were located in the northern region, 9 in the middle region, and 6 in the southern region. Schools varied in their identity: 14 schools were grounded in a religious denomination and 9 schools applied innovative teaching concepts, such as Montessori or Dalton. There were 20 fourth grade classes, 14 fifth grade classes, 20 sixth grade classes, and 11 multigrade classes combining two or three grades participating in the experiment.

Teachers, volunteering to participate in the study (N = 30), were assigned to the teacher-trainer group. This group was provided a professional development program in writing education by the researchers. These teachers had to bring at least one colleague who they trained themselves for the duration of the study. So, the sample included at least one teacher-trainer and one teacher-trainee from each school. In total, there were 38 teachers in the teacher-trainee group. Teacher-trainers had slightly more teaching experience than teacher-trainees (M = 13.83 years, SD = 8.20 versus M = 8.98 years, SD = 8.21, $t(63) = 2.38$, $p < .05$), but they did not differ significantly in years of experience

in the grade in which they were currently teaching (M = 3.51, SD = 3.59, $t(63) = 1.51$, p = .14).

Table 1 presents the number of students per grade in each group. The trainer group consisted of 28 classes with a total of 602 students. The trainee group consisted of 37 classes with a total of 763 students. As can be seen in Table 1, there were no differences between the two groups in the percentage of female and male students ($\chi^2(2) = .25$, $p = .62$). Students' age ranged from 8 to 13 years, with an average of 10.23 years (SD = 1.00), which did not differ between groups ($F(2) = 0.81$, $p = .37$). There were minor differences in the language background of the students of both groups ($\chi^2(2) = 16.73$, $p < .001$). Dutch was the native language for most students in the trainer (64%) and trainee group (74%), but the trainer group consisted of relatively more students for which Dutch was the L2 (36%) than the trainee group (26%). The most frequently spoken languages besides Dutch were Arabic, Turkish, English and Frisian (a language spoken in the northern regions of the Netherlands).

In total there were 19 students who dropped out during the study because they changed schools: 11 students in the trainer group and 8 students in the trainee group. These students were removed from the data set, which resulted in a total sample of 1346 students.

2.2. Design

The intervention in this study is implemented by using a design with switching panels (Shadish, Cook, & Campbell, 2002), with two groups (trainers and trainees) and three measurement occasions. In this design the intervention is implemented in both groups, but at different moments in time. An advantage of this design is that, as the intervention is implemented consecutively in the two groups, it is possible to test whether

the effectiveness of the intervention differs between teacher-trainers and teacher-trainees. If the effect of the intervention is equal for both groups, this indicates that professional development provided by experts or colleagues is equally effective, meaning that professional development can be transferred between colleagues and does not rely on experts. Another advantage of a switching panel design is that all students eventually benefit from the intervention, making it a more ethical design than a regular pre-post (quasi-) experimental design, see also Bouwer et al, (2016a).

Teachers and students in the trainer group started with the intervention program in the first period, between the first and second measurement occasion. This period lasted four months during which teachers executed one writing lesson a week instead of their regular program for writing. While the trainer group started with the intervention, the trainee group served as a control group, engaging in their regular writing activities and routines. Whereas the second measurement occasion for the trainer group was scheduled after four months (after completing the whole intervention program), it was scheduled already after two months for the trainee group. This overlap in time was created for an optimal implementation of the teachers-training-teachers approach, as jointly working on the program would promote collaboration and interaction between trainers and trainees (Borko, 2004).

After the second measurement occasion, teachers and students in the trainee group started with the intervention, while teachers and students in the trainer group returned to their regular writing activities. The procedure for the trainee group was the same, teachers executed one *Tekster*-lesson a week for a period of four months. The only difference between the two groups was the professional development of teachers. Teachers in the trainer group received two expert training sessions (before and during the intervention program), whereas teachers in the trainee group were trained by their expert-

trained colleagues. The third measurement occasion served as a posttest for students in the trainee group, as well as a delayed posttest for students in the trainer group, with which we were able to measure retention.

2.3. Regular writing education

In the present study, the intervention program is compared to the regular classroom practice in writing education. In the Netherlands, writing education is traditionally part of the language teaching curriculum. A report from the Dutch Inspectorate of Education (Henkens, 2010) showed that from the 8 hours per week reserved for language teaching, on average 45 minutes are spent on writing. These writing lessons are primarily product-focused: students hardly receive any support during the writing process, nor are they supported on how to approach writing tasks. In the majority of schools the writing performance of students is not monitored, and students are seldom given feedback on their performance. Questionnaire data on the regular classroom practice of the teachers participating in the present study show comparable results. A majority of teachers (70%) indicated to use textbooks for language teaching for their writing lessons. On average, they devoted 42.42 minutes (SD = 29.91) a week to writing in class, of which 13.93 minutes (SD = 12.11) were devoted to instruction in writing strategies. Teachers provided their instruction mostly plenary, less time was devoted to small-group instruction or individualized instruction. Overall, there were no differences between trainers and trainees in their regular writing practice (p > .06).

2.4. Writing intervention: Tekster

The intervention program consisted of a teaching program, *Tekster*, which included three lesson series of 16 lessons, one for each grade level, compiled in a

workbook for students, accompanied by a teacher manual (Koster et al, 2014a, 2014b, 2014c). Further, to foster the professional development of teachers and support them in the implementation of the program two additional training sessions were provided. In *Tekster* the main focus of instruction was to teach students a strategy for writing, supplemented with explicit instruction in text structure and the teaching of self-regulatory skills. To support students in applying the writing strategy, they were taught a mnemonic representing the steps of the writing process: VOS *(fox)* for grade 4, DODO *(dodo)* for grade 5, and EKSTER *(magpie)* for grade 6. The letters of the acronyms represent the steps in the writing process as follows: VOS (fox) for Verzinnen (generate content), Ordenen (organize), Schrijven (write); DODO (dodo) for Denken (think), Ordenen (organize), Doen (do), Overlezen (read); Ekster (magpie) for Eerst nadenken (think first), Kiezen & ordenen (choose & organize), Schrijven (write), Teruglezen (reread), Evalueren (evaluate), Reviseren (revise). In the first lesson of the program the acronym-animal was introduced in a story in which students also practice the steps of the strategy for the first time. In the following lessons the animals are the common theme, with small icons of the animals serving as a visual support.

Table 2 gives a general description of the design principles of the program and how these principles were operationalized in learning and teaching activities (see Rijlaarsdam, Janssen, Rietdijk, & Van Weijen, 2015). In particular it shows how effective practices for mode and focus of instruction are combined and translated into learning activities for students and teaching activities for teachers, as well as the support that is provided for teachers in the teacher manual and during the training sessions of the professional development program.

2.4.1. General lesson format

To ensure that all activities described in Table 2 were covered, all *Tekster*-lessons were designed using a general (more or less fixed) lesson format. The core of the lessons was the overall writing strategy. Each lesson started with a plenary introduction in which the goal of the lesson is explicitly stated. Specific characteristics of the text type were addressed through modeling (teacher modeling, or peer modeling using videoclips), comparing model texts, or explicit instruction. Next, the authentic writing assignment was introduced, with explanation of the communicative goal and intended audience, and the acronym for the strategy was explicitly named. Subsequently, students started with the first step of the strategy, which was generating content in keywords, followed by the second step, which was organizing the generated content, supported by the teacher through scaffolding with gradual release of responsibility. In the third step of the lesson students started writing their texts using the organized content, while the teacher provided support when necessary. In the following step (grade 5 and 6 only) students read each other's texts or their own text. In the fifth step of the lesson (grade 6 only) students evaluated the written text by answering evaluative questions and/or giving feedback. In the sixth step of the lesson (grade 6 only), students revised (parts of) their text on the basis of the feedback they received. The duration of the average *Tekster*-lesson was between 45 and 60 minutes. A sample lesson is included in Appendix A.

2.4.2. Writing tasks in the lesson program

In the program students learn to apply the writing strategy to various types of texts, for which authentic writing tasks with various communicative goals and audiences are used. For instance, in each grade they learn to write descriptive texts (e.g., a self-portrait or personal ad), narrative texts (e.g., a story or newspaper article), persuasive texts (e.g., a nomination email for a television program or a flyer for recruiting new

members for a club), instructive texts (e.g., a recipe, rules for a game) and personal communication (e.g., a holiday postcard or invitation). The writing tasks comply with the goals set for the end of elementary school by the Ministry of Education.

The level of difficulty ascended through the grades as follows: in grade 4, predominantly writing tasks were used in which the intended audience was in close proximity of the student, such as classmates, friends, or (grand-) parents. In grade 5, this was expanded to people with whom students have a more distal relationship, but are still familiar to them, such as their teacher, relatives, or neighbours. In grade 6, students also have to write texts that are intended for unfamiliar people, such as the editor of a newspaper, or the managing director of a company. The writing tasks were developed in close collaboration with elementary teachers to ensure that the topics would match students' interest and developmental level. The teachers piloted the writing tasks in their own classrooms first, and subsequently, the program was tested in a pilot study (Koster, Bouwer, & Van den Bergh, 2016) and in a large-scale intervention study (Bouwer et al, 2016a).

2.4.3. Professional development

The professional development component of the program consisted of a teacher manual and two training sessions.

2.4.3.1. Teacher manual. The teacher manual was provided to all teachers in order to facilitate the teaching of *Tekster*-lessons. The manual consisted of two parts: a general introduction and detailed lesson plans for each lesson. In the general introduction the goal and approach of the program were explicated as well as the general lesson format. Further, the importance of feedback for learning to write was explained and suggestions were given on how to provide effective feedback. Next, observational learning and

modeling were explained, and practical information was given concerning the organization of the intervention. The detailed lesson plans provided an overview of the instruction and activities of the lesson with a time planning for each phase of the lesson. These lesson plans described the activities the teacher was expected to execute during the lesson and provided suggestions when to use modeling during instruction. The manual also included a dvd with movie clips of peer modeling to use during instruction, and videos with examples of teacher modeling for different phases of the lesson. Further, the manual included a benchmark rating scale to support teachers in evaluating text quality and giving feedback. The benchmark rating scale consisted of five students' texts of ascending quality, representative of the range of text quality that can be found in grade 4 to 6 (Bouwer, Koster, & Van den Bergh, 2016b).

2.4.3.2. Training sessions. To support teachers in the implementation of the program in the classroom, two training sessions were planned over the course of the intervention. During these sessions, which lasted four hours each, teachers were trained in small groups (max. 12 teachers) by the researchers. The righthand column in Table 2 demonstrates how the aspects of the training are related to the learning activities and the teaching activities of the program. During the first training session prior to the start of the intervention teachers were briefed on how to work with the program. The researchers informed the teachers about the theoretical background, and showed and discussed an example video of teacher modeling. Teachers were instructed to apply coping modeling during their lessons. In contrast to mastery models, who show a flawless performance, coping models initially display exemplary deficiencies but overcome these difficulties and gradually improve their performance (Schunk, 1987; Zimmerman & Kitsantas, 2002). Research has shown that observing coping models raised students' self-efficacy and enhanced their performance more effectively than a mastery model (Zimmerman and

Kitsantas, 2002). This may be due to the explicit modeling of strategies to overcome difficulties, or it might be that, due to perceived similarity to the model, students believe that they are also able to improve their performance (Schunk, 1987). During the training session was discussed how teachers could implement modeling in their own writing lessons. Next, the teachers, in small groups, jointly prepared the first two lessons with special attention to where and how modeling could be applied in these lessons. Lastly, teachers were instructed to read the information in the manual carefully and watch the videos before the start of the program as a preparation for teaching the lessons.

The second training session was scheduled after six lessons. During this session, first experiences and specific issues regarding the implementation of *Tekster* in the classroom were shared and discussed. Next, teachers were trained in how to provide their students with effective feedback, and how to assess students' texts. For effective feedback teachers have to adjust their comments to students' needs, which requires that teachers are able to assess the quality of students' texts and adapt their feedback accordingly (Bouwer, Koster, & Van den Bergh, 2016c). Therefore teachers were trained in how to evaluate text quality using benchmark texts representing ascending levels of writing quality (Bouwer, et al, 2016b). After this, they were introduced to the characteristics of effective feedback, followed by a plenary discussion of examples of teacher feedback. Subsequently they practiced how to provide effective feedback using the scale with benchmark texts: with example texts, but also with texts of their own students. They reflected on the quality of their feedback in subgroups.

2.4.3.3. Teachers-training-teachers approach. For the training sessions we adopted a teachers-training-teachers approach: teachers in the first group were trained by the researchers, subsequently these teachers trained their colleagues. Teachers who followed the professional development program received instruction and materials to

subsequently train their colleagues, who started with the program two months later. The teacher-trainers were instructed to plan two training sessions with their colleagues, in which the same topics should be addressed as in the training sessions that they received. Further, it was encouraged that teacher-trainers would invite their colleagues in their classroom to observe a *Tekster*-lesson.

2.5. Measures and procedure

2.5.1. Assessment of students' writing quality

To examine how the intervention program affected students' writing performance, students completed three writing tasks prompting for different genres: descriptives, narratives and persuasive letters at each measurement occasion. The tasks within a genre were similar with regards to the communicative goal and intended audience, and only differed by topic. Hence, students wrote nine texts in different genres and topics, which warrants generalization to writing proficiency (Bouwer, Béguin, Sanders, & Van den Bergh, 2015). Each task contained a writing prompt, including an illustration with relevance to the topic, and some space for prewriting which students were free to use. Appendix B provides examples of writing prompts for each genre. Similar writing tasks were used and validated in a previous study with students in the same age group (Bouwer et al, 2016a). Teachers administered the writing tasks to their students during normal class time, without providing any additional instruction. Students had to work individually on the task, without a time limit. Teachers were instructed to plan the three writing tasks for each measurement occasion within one week, but not on the same day.

The quality of students' texts was rated by eighteen experienced elementary teachers using a benchmark rating procedure. In this procedure, raters independently score each text by comparing it to a scale with five benchmark texts (Bouwer et al,

2016b). These benchmarks reflect the range of writing quality of students in grade 4 to 6. There were different benchmark scales for each writing genre, see Appendix C for an example. The center position on each scale is an average text which is assigned an arbitrary score of 100 points. The other texts on the scale are one (115 points) and two (130 points) standard deviations above average, and one (85 points) and two (70 points) standard deviations below average. This rating procedure was developed in a previous study in which its support for raters in assessing text quality across tasks and genres was demonstrated (Bouwer et al, 2016a, 2016b). To ensure that raters were blind to conditions, we anonymized students' texts. Each text was rated by a jury of three raters using a design of overlapping rater teams (Van den Bergh & Eiting, 1989). In this design, texts are randomly divided into subsamples, equaling the number of raters (N = 18). Subsequently, each rater received 3 subsamples according to a prefixed design. Because each subsample was rated by overlapping rater teams, it was possible to estimate the reliability of the scores of individual raters, and to approximate the reliability of jury raters (Van den Bergh & Eiting, 1989). The average reliability of jury ratings was high in the present study, overall ρ = .88, varying from ρ = .83 to ρ = .90 per task. The final text quality score was determined by averaging the scores of the jury raters. As scores appeared to be somewhat negatively skewed, raters' scores were normalized for each task using Blom's rank-based normalization formula (Solomon & Sawilowsky, 2009).

2.5.2. Teachers' self-efficacy and attitudes for writing

To gain insight into the influence of the intervention program and the professional development program on teachers' self-efficacy, attitudes towards writing and the teaching of writing, teachers filled in questionnaires prior and after the intervention program. Teacher efficacy for writing was measured by the Efficacy Scale for Writing

(TES-W; Brindle, 2013; Graham et al., 2001). This scale measured teachers' beliefs about their own writing instruction on two dimensions: (1) the degree to which teachers attribute students' successful writing to their own writing instruction (3 items, e.g., "When students' writing performance improves, it is usually because I found better ways of teaching them") and (2) the perception of their ability to support inexperienced writers (4 items, e.g., "When I try really hard, I can help students with the most difficult writing problems"). The items were measured on a 5-point Likert scale ranging from 'strongly disagree' to 'strongly agree'. The scale has been validated in previous research (Brindle, 2013; De Smedt et al., 2016) and the internal consistencies for the subscales in the present study were satisfactory, respectively α = .77 and α = .58. Positive medium correlations between the subscales on pretest measures confirmed that the scales are related but measure different dimensions of teacher efficacy (r = .39, p < .01).

Teachers' attitudes for writing were measured by the questionnaire of Brindle (2013). The questionnaire included 4 items on teachers' attitudes towards writing (e.g., "I enjoy writing") and 4 items on their attitudes towards writing instruction (e.g., "I like to teach writing"), which were measured on a 5-point Likert scale ranging from 'strongly disagree' to 'strongly agree'. The scale has been validated in previous research (Brindle, 2013; De Smedt et al., 2016) and the internal consistencies for the subscales in the present study were high, respectively α = .84 and α = .89. Positive medium correlations between the subscales on pretest measures confirmed that the scales are related but measure different dimensions of teacher attitudes (r = .26, p < .01).

2.5.3. Implementation and social validity of the intervention program

We used an explanatory sequential mixed methods design (Creswell & Plano Clark, 2011; Johnson, Onwuegbuzie, & Turner, 2007) to examine whether trainers and

trainees implemented the lesson program as intended, to investigate how trainers implemented the professional development program, and to get in-depth information about the social validity of the intervention. In the first phase we collected and analyzed quantitative data from logbooks, post-intervention questionnaires, and observations. The triangulation of the data of different measures provides a better understanding than either approach alone. Whereas questionnaires and logbooks provide general information on teachers' classroom practice and experiences, observations provide richer and more objective data about how teachers actually implemented the program (Desimone, 2009). In the second phase we collected qualitative data from focus group interviews. By exploring differences between trainers and trainees in more depth, the qualitative data further refine and explain the quantitative results obtained in the first phase (Creswell & Plano Clark, 2011). The specific qualitative and quantitative measures and procedures are further explained below.

2.5.3.1. Logbooks. Logbooks were incorporated in the teacher manual. Teachers were requested to fill in the logbook after each lesson. They were asked to provide the following information about each lesson in the program: preparation time, lesson duration, appreciation of the lesson (on a scale from 1 to 10; 1=very low, 10=very high), estimated level of difficulty for their students (on a 5-point scale; 1=easy, 5=hard), the level of difficulty of executing the lessons (on a 5-point scale; 1=easy, 5=hard), and additional comments (if any). After the intervention the logbooks were collected, together with the students' workbooks, to check whether all writing lessons were executed as planned.

2.5.3.2. Observations. Classroom observations generally provide more insight in the actual implementation of an intervention (Desimone, 2009). To obtain reliable and valid information on the actual implementation of the program in the classroom,

observations of multiple lessons over an extended period of time are required (Desimone, 2009). Therefore we observed three lessons in the classrooms of a sample of the participating schools. We used a heterogeneity sampling method to get a representative view insight on how *Tekster* is implemented in different contexts. Schools were selected based on four criteria: (1) geographical location (urban, suburban or rural), (2) student population (predominantly students with Dutch as their first language (L1) or predominantly students with Dutch as their second language (L2)), (3) classroom organization (single or multi-grade classrooms), and (4) teachers' appointment (fulltime, with one teacher per classroom, or parttime, with two teachers per classroom). This selection procedure resulted in a sample of five schools, see Table 3. This subsample included one-third of the total sample of teachers, with 10 teacher-trainers and 12 trainees from 17 classes.

During the intervention period we observed multiple lessons in each of the 17 classrooms. We observed lessons of trainers as well as trainees to investigate whether classroom practice differed between these groups. In total, we observed 24 lessons of trainers and 30 lessons of trainees. We used an observation instrument developed for a previous study in which the effectiveness of *Tekster* was examined (Bouwer, 2016a). The instrument consisted of two parts for each phase of the lesson. In the first part was tallied every 20 seconds whether the teacher was on task or off task: on task if the teacher was executing the actions as specified in the lesson plan for that particular phase of the lesson, off task if the teacher was involved in other activities than teaching writing, such as fetching a cup of coffee or talking to a colleague. Further, it was tallied whether the on-task-behavior involved plenary activities (instruction or classroom interaction) or interaction with individual students. Second, for every lesson phase observers had to register whether the teacher applied key components of the intervention, i.e., teacher

modeling, referring to the acronym and/or the steps of the strategy, or providing feedback. Additionally, the time for each phase of the lesson and for the lesson as a whole was registered. Each classroom was observed by one trained undergraduate student, there were ten observers in total. To optimize observers' agreement, all observers were trained in advance.

2.5.3.3. Post-intervention questionnaire: Implementation of the lessons. To measure teachers' implementation of the lesson program *Tekster*, teachers were first asked to indicate whether they adapted lessons to their own context on an ordinal scale with three categories (never, sometimes, always), and to provide reasons for this. The next nine items measured how teachers implemented the key components modeling, feedback, and evaluating text quality. Three items measured how often during the intervention teachers modeled (a part of) the writing process, how often they provided feedback to students, and how often they evaluated students' writing products. This was measured on an ordinal scale with five categories: never, only once, in some lessons, every lesson, multiple times per lesson. Three items measured whether teachers used each component more frequently during the intervention than during their regular classroom practice for writing on a scale from 1 (less frequently) to 3 (more frequently). For the component feedback, they also had to indicate whether their feedback focused mainly on the writing product, the process, or both, and whether they provided mainly oral or written feedback. Further, they were asked to indicate whether they used a benchmark rating scale for evaluating text quality, providing feedback, and/or instruction in class.

2.5.3.3. *Post-intervention questionnaire: Implementation of the training.* To get more insight in whether the collegial training sessions were comparable to the expert training sessions, the questionnaire included five items measuring teachers' implementation of the teachers-training-teachers approach. Teachers were asked how

often they organized a training session with their colleague(s), which was measured on an ordinal scale with four categories (never, once, twice, or more than twice), and they had to indicate the total training time in minutes. Further, they were asked whether the training was one-on-one or in a team setting. To get more information on the content of the collegial training sessions, we asked them to indicate whether they discussed the following topics: the goal and structure of the program, organization of the lessons, modeling, feedback, rating text quality, the benchmark rating scale, student texts or specific issues. They also had to indicate whether they observed a *Tekster*-lesson of a colleague.

2.5.3.4. *Post-intervention questionnaire: Social validity of the intervention program.* The questionnaire further measured the social validity of the intervention program, including the lesson program and professional development activities. First, to measure teachers' satisfaction with the lesson program, trainers and trainees had to indicate their general attitude towards the lesson program *Tekster* on a 5-point Likert scale from 'highly negative' to 'highly positive'. They were also asked to indicate on a 5-point Likert scale from 'strongly disagree' to 'strongly agree' whether they believed that their students were supported by key components of the intervention, i.e., writing strategy, teacher modeling, peer modeling clips, model texts, feedback, and peer interaction. Further, they had to indicate whether they intended (or not) to keep using *Tekster*, components or specific lessons, after the study. Finally, to gain insight into students' satisfaction with the lesson program, we asked students to rate the overall program on a 10-point scale from 'very bad' to 'very good'. Teachers' experiences with the professional development activities were measured with six questions. Teachers had to indicate on a 5-point Likert scale whether their classroom practice was supported by the teacher manual, and by the benchmark rating scale that was included in the manual.

Further, teachers had to indicate on a 5-point Likert scale whether the training sessions provided general support for their writing instruction, and specifically for modeling, providing feedback, and evaluating the quality of their students' texts. Together these questions provided information on the usability of *Tekster* in daily classroom practice and whether the knowledge and skills required for using *Tekster* effectively were easily transferable between colleagues.

2.5.3.6. Focus group interviews. To get more in-depth information on differences between trainers and trainees in how they implemented and experienced the intervention program including professional development activities, we conducted semi-structured focus group interviews after the intervention period with all participating teachers at the schools where we observed during the intervention period. Hence, we conducted five interviews at five different schools, with 9 trainers and 11 trainees in total (unfortunately, 1 trainer and 1 trainee were not able to participate). The interview protocol focused on three main themes: the lesson program *Tekster*, the teacher manual and the training sessions. The results from the quantitative data were used as input for these themes, in order to further explain differences between trainers and trainees and to explore possible factors that might have affected intervention fidelity.

The protocol began by asking trainers about their experiences with the implementation of the program in their classroom, differentiation between weak and strong writers, the content of the lessons, and the general format of the lessons, i.e., the steps of the writing strategy. We used open, non-directive questions to limit socially desirable responses and followed up with 'why' and 'how' clarification questions until we reached full understanding. After that, we asked trainees the same questions, using the same procedure. When both trainers and trainees did not have anything to add to this theme, we continued with the second theme, the teacher manual. For this theme, we

asked teachers to indicate whether they used the general introduction, the lesson plans, and the benchmark rating scale in the teacher manual for preparing their writing lessons, and what their opinion was on these aspects of the manual. We applied the same procedure: first we asked the trainers, then the trainees. The same procedure was followed for the third theme, the training sessions, in which we asked both trainers and trainees whether they experienced enough support to execute the lessons. To get more information on the content of the teacher-training sessions, we asked trainers what aspects of the training program they transferred to their colleagues and whether they experienced enough support to do this. Trainees were asked to indicate how often they experienced some sort of support from their trainers and what kind of support that was. Finally, teachers were offered the possibility to offer recommendations for improving the lesson program *Teksters* based upon their experiences.

Interviews were conducted by two interviewers: one asking the main questions, the other taking notes and summarizing what was said during the interview in order to enhance the reliability and validity of the interpretation of the interview data. The interviews were audiotaped and lasted from 50 to 67 minutes in duration. Afterwards they were transcribed and coded by an independent coder, an undergraduate student who was trained in advance on the core features of the intervention program.

2.6. Data analysis

2.6.1. Students' writing quality

Scores for students' text quality were hierarchically organized; scores were cross-classified with students and tasks, and students were nested within classes. The data are therefore analyzed by applying six multilevel models in which parameters were added systematically, in order to test the effectiveness of the intervention program and to test

whether there are differences between the trainer and trainee group. In this type of models all students, including those with partly missing values, are taken into account. Inspection of the data revealed that only 4% of the students missed four or more of the writing tasks.

Model 1 is the basic null model in which we only account for random error (S^2_e) and random effects of students (S^2_s), tasks (S^2_t), and classes (S^2_c). That is, writing scores are allowed to vary within and between students, between tasks (including systematic variation due to genre), and between classes. In Model 2 training group is added as a fixed effect to test whether average scores differ between trainers and trainees. In Model 3 measurement occasion is added as a fixed effect to test whether average scores differ over time. Model 4 tests the main effect of the writing intervention by estimating the interaction between group and measurement occasion. As trainers and trainees received the intervention at different time intervals, we tested the interaction by comparing the slope of the regression line of students' writing scores in the trainer group between the first and second measurement occasion with the slope of the regression line of students' writing scores in the trainee group between the second and third measurement occasion (instead of comparing average scores on the pre- and post measures). This model includes the restriction that the effect of the intervention is the same in the two groups. In Model 5 this restriction is removed to test whether the effectiveness of the intervention differs between the trainer and trainee group. In Model 6 grade is added as a fixed effect to test whether average scores differ between grades.

2.6.2. Teachers' self-efficacy and attitudes for writing

To analyze how the intervention including the professional development program affected both trainers and trainees, we analyzed differences between teacher-trainers and

teacher-trainees on self-efficacy, and attitudes for writing and teaching writing. We used a repeated measures MANOVA to determine whether teachers' feelings of self-efficacy and attitudes for writing and teaching writing changed over time, due to following the intervention program. We also tested whether there were differences between trainers and trainees, and whether there were differences between grades.

2.6.3. Implementation and social validity of the intervention

Teachers' implementation of the intervention program was measured quantitatively by logbook entries after each lesson, questions in the post-intervention questionnaire, and classroom observations. We first analyzed these measures separately. For each measure we analyzed whether there were significant differences between trainers and trainees in how they implemented the lesson program or key components of the program. Next, we triangulated the quantitative data in order to check whether they converged or diverged.

Next, we coded the qualitative data from the focus-group interviews. For each interview separately, (sub)categories were first identified and coded. Subsequently, the subcategories of all interviews were compared in order to identify important themes and subcategories of information across the five interviews. The themes (and subcategories) that were identified from the transcripts were comparable to the themes (and subcategories) from the interview protocol, i.e. respectively, the lessons (implementation, differentiation, content, writing strategy), the teacher manual (general introduction, lesson plans, benchmark rating scale) and the training sessions (support from training, collegial training, content of training). We also marked specific quotes that served as a clear illustration of the views of trainers and trainees.

In a following step, at the interpretative level of research, we linked the quantitative data from logbooks, questionnaires and observations to the qualitative data from the focus-group interviews, to compare and explain the most prominent findings. Integrating these two research methods provides a more complete understanding of how teachers implemented and experienced the program than either set of methods on its own (cf. Yoshikawa, Weisner, Kalil, & Way, 2008). For instance, differences between self-reported data (i.e., what teachers say they do) and observational data (i.e., what teachers actually do) can be explained by qualitative data from focus-group interviews.

3. Results

3.1. Effect of the intervention on students' writing quality

We first tested the effect of the intervention on students' writing performance. Table 4 shows the results of the fit and comparison of the planned models. As can be seen, there was no main effect of training group (Model 2 versus Model 1, $\chi^2(1) = 1.74$, $p = .19$), indicating that average writing scores did not differ between students from teachers who were trained by experts and students from teachers who were trained by their colleagues. There was, however, an effect of measurement occasion (Model 3 versus Model 1, $\chi^2(3) = 101.61$, $p < .001$), indicating that scores were not the same on the three measurement occasions. Results also showed an interaction effect between training group and measurement occasion (Model 4 versus Model 3, $\chi^2(2) = 4564.73$, $p < .001$), indicating that differences in writing scores between two measurement occasions (i.e., between occasion one and two, or between occasion two and three) were not the same for students in the trainer and trainee group.

With the fifth model we tested whether the effectiveness of the intervention (indicated by an interaction between training group and measurement occasion) was

different between the trainer and trainee group. This model was not significant (Model 5 versus Model 4, $\chi^2(1) = 1.50$, $p = .22$), indicating that there were no differences between trainers and trainees in the effectiveness of the intervention.

Parameter estimates of the fourth model are summarized in Table 5. While generalizing over students, teachers, and tasks, students' writing improved with 7.29 points (*SE* = 0.58, *t* = 12.57, *p* < .001), which was a medium effect (Cohens *d* = 0.55). Between the first two measurement occasions, students in the trainee group served as a control group. It can be seen that their writing performance did not improve during this period (β = 0.31, *SE* = 0.60, *t* = 0.52, *p* = .30). Further, the effect of the intervention maintained over time: there were no significant differences in writing scores between the posttest directly after the intervention and the delayed posttest after two months in which students in the trainer group returned to their regular writing program (β = 0.36, *SE* = 0.64, *t* = 0.56, *p* < .29). A graphical overview of the effect of the intervention in both conditions is presented in Figure 1.

Table 4 further shows a main effect of grade (Model 6 versus Model 4, $\chi^2(2) = 33$, $p < .001$), indicating that average writing scores were different for students in grade 4 to 6. On average, grade 5 students scored 5.65 points (*SE* = 1.39) higher than students in grade 4. The scores of students in grade 6 were even higher: they scored 8.69 points (*SE* = 1.39) higher than the students in grade 4. Hence, the average improvement in writing quality per grade yielded 4.78 points. If we compare the improvement due to the intervention to the general improvement in writing skills of students between grade 4 to 6, the magnitude of the effect of the intervention becomes even more visible. This comparison shows that students' writing improved by more than one-and-a-half grade after following the four-month writing intervention program.

We approximated the magnitude of the effect of the intervention on students' writing performance, for the whole intervention program consisting of a total of 16 lessons. Inspection of students' workbooks revealed, however, that not all teachers managed to complete the whole intervention program. On average, they completed 13 lessons (SD = 2.80). This affected students' writing scores: for each missing lesson, the effect of the intervention decreased with 0.58 points (SE = 0.12, t = 5.17, $p < .001$). There was no significant difference between the trainer and trainee group ($\chi^2(1) = 0.43$, p = .51).

3.2. Effect of the intervention on teachers' self-efficacy and attitudes

The significant improvements in students' writing performance after the intervention indicated that trainers as well as trainees effectively implemented the writing program in their classrooms. To further analyze how the intervention including the professional development program affected teachers, trainers as well as trainees, we analyzed differences between teacher-trainers and teacher-trainees on self-efficacy and attitudes for teaching writing. Multivariate test results of the repeated measures MANOVA indicate that for at least one of the dependent variables there was a significant main effect of training group ($F(4, 49) = 3.31$, $p < .05$, $eta^2 = .21$), and time ($F(4, 49) = 4.05$, $p < .01$, $eta^2 = .25$). There was no main effect of grade ($F(12, 129.93) = 0.93$, p = .52). Table 6 shows the average scores for teachers' self-efficacy and attitudes toward writing and teaching writing over time. Univariate test results indicate that teachers were more positive about teaching writing after the intervention than before ($F(1, 52) = 5.99$, $p < .05$, $eta^2 = .10$), they felt more efficacious to teach successful writing ($F(1, 52) = 8.10$, $p < .01$, $eta^2 = .14$), and they felt more efficacious to support inexperienced writers ($F(1, 52) = 10.52$, $p < .01$, $eta^2 = .17$). Their general attitudes towards writing remained the

same ($F(1, 52) = 1.15$, $p = .29$). Although the average scores for teachers' self-efficacy for successful writing and their attitudes towards teaching writing seemed to be slightly higher for trainers than for trainees (respectively, $F(1, 52) = 8.11$, $p < .01$, $eta^2 = .14$ and $F(1, 52) = 8.42$, $p < .01$, $eta^2 = .14$), there were no significant interaction effects between time and group for any of the subscales ($F(1, 52) < 1.56$, $p > .22$). Taken together, the results indicate that the intervention had the same effect on trainers and trainees: both groups of teachers became more positive and felt more efficacious about the teaching of writing.

3.3. Teachers' classroom implementation of the intervention

3.3.1. Implementation of the lesson program

The logbook and questionnaire data, together with the classroom observations provided information on how teachers implemented the lessons in the classrooms. In total, 87% of the logbooks were filled in and returned. The logbook data show that the average preparation time for the lessons was 12 minutes ($SD = 6.4$), which did not differ between trainers and trainees ($F(1, 93) = 1.28$, $p = .26$). The average lesson time as reported by teachers in the logbook was 46 minutes ($SD = 6.98$), which was according to the planning in the lesson plans in the teacher manual and did not differ significantly between trainers and trainees ($F(1, 93) = 3.28$, $p = .07$). The lesson time reported by the teachers converged with the observation data, which showed that the duration of the average observed *Tekster*-lesson was 43 minutes. However, not all lessons were fully completed in this time: in 28% of the grade 5 and 6 lessons the post-writing phase (reread, evaluate and revise) was not covered during the observed lessons, due to time constraints or differences in pace between students. Teachers indicated that these components would be accomplished later in the week, during students' independent work

time. The observational data also revealed that teachers were on task on average 90% (*SD* = 10) of the lesson time, which means that they were engaging in the activities as prescribed in the lesson plan, either plenary with all students (especially during the first phases of the lessons), or with individual students. There were no significant difference between trainers and trainees in on-task behavior ($F(1, 52) = 0.70$, $p = .41$). Questionnaire data revealed that most of the teachers (86%) adjusted some or all lessons to their own context. There were no differences between trainers and trainees ($\chi^2(2) = 0.92$, $p = .63$). Main reasons for adjustments were: classrooms combining multiple grades (combining lessons from different workbooks), students with NT2 (additional language support), or including examples from real-life, the news or students' own experiences.

Together, the quantitative findings showed that teachers implemented the content and structure of the lessons as intended, but they struggled to finish all lessons in the planned time. The qualitative findings from the interviews confirmed these quantitative findings. During the interviews, both trainers and trainees indicated that for a number of lessons the planning of the lessons was too tight, and that they did not succeed in completing all lesson components during the planned lesson time. This was especially the case for grade 5 and 6 teachers, for whom the lessons included more post-writing activities. Not all students were able to finish their texts at the same time, which complicated the organization of post-writing activities. As suggested in the manual, most of the teachers divided the lessons in two parts. In the first part students planned and organized ideas and started with writing their text. In the second part students evaluated and revised their text. Between the first and second part students were able to finish their text at their own pace. A very positive organizational point according to the teachers was that they only needed little time to prepare the lessons: only 10 minutes on average as indicated by the questionnaire data. In the interviews it became clear that most teachers

spent even less than 10 minutes on preparing their lessons. They indicated that the lesson plans and the goals of each lesson in the manual were very clear. They further indicated that preparation was facilitated by the well-defined overall structure of the lessons: for each lesson, regardless writing genre, students had to apply the same steps of the writing strategy. Some teachers reported that they adjusted the content of writing lessons to fit students context and/or experience even more. For example, in one of lessons students had to write an invitation for the end-of -year musical. One of the teachers had her students write an invitation for their own musical, based on this writing task. Another teacher collected recent examples from the news and internet for text structure instruction instead of the examples provided in the lesson material. Teachers from multiple grade classrooms indicated that during their lesson preparation they paid specific attention to how to combine the lessons for students in different grades into one writing lesson for the whole class. This was especially challenging for lessons in which the communicative goal of the lesson was not the same across grades. Furthermore, one trainee indicated that the preparation depended on her familiarity with the text genre: For more unfamiliar genres, she had to collect additional background information and example texts to increase her own knowledge about the particular genre.

3.3.2. Implementation of key components

Regarding the implementation of key components of the intervention program, i.e., modeling, strategy use, feedback, and evaluating text quality, questionnaire data showed that all teachers modeled (a part of) the writing process during the intervention: 57% of the teachers modeled in some lessons, 35% of the teachers modeled in every lesson, and 8% of the teachers indicated to model multiple times in every lesson. One-third indicated that this was more frequent than during their regular writing instruction,

half of them indicated that it was the same. For feedback a similar result was found: 62% of the teachers indicated to have provided feedback in some of the lessons, 27% provided feedback in every lesson, 10% provided feedback multiple times in every lesson, and only 2% provided feedback only once. Again, one-third of the teachers indicated that this was more frequent than in their regular practice, and half them indicated that it was the same. The majority of the teachers provided both process- and product-related feedback, and most teachers provided a combination of oral and written feedback. There were no differences between trainers and trainees in how often they implemented modeling and feedback in their lessons, $\chi^2 < 4.39$, $p > .11$.

The questionnaire data converged with the observational data, which indicated that both trainers and trainees implemented the key components of the intervention in their classrooms to an equal extent. On average, they modeled once during each lesson ($M = 1.00$, $SD = 1.32$). On average, they explicated the acronym once ($M = 1.22$, $SD = 1.20$), and the strategy twice ($M = 2.13$, $SD = 1.55$) during the lessons. Feedback was provided multiple times ($M = 2.76$, $SD = 1.54$). The results of the observational data suggest that teachers executed the lessons as planned in the lesson plans in the manual and there were no differences between trainers and trainees in the observed components, $F(1, 52) < 2.80$, $p > .10$. There were, however, differences between trainers and trainees, in how often they evaluated their students' written products ($\chi^2(4) = 16.46$, $p < .01$). The questionnaire data showed that the majority of the trainers did this only once (63%) or in some lessons (35%), whereas the majority of trainees evaluated students' texts in some lessons (47%) or in every lesson (29%). Whereas 21% of the trainers indicated that they evaluated text quality more frequently than in their regular practice, this was only found for 9% of the trainees ($\chi^2(2) = 6.07$, $p < .05$). Results also showed that trainers and trainees differed in the use of the benchmark rating scale for evaluating text quality (93%

versus 74%, $\chi^2(1) = 4.16$, $p < .05$) or providing feedback (86% versus 62%, $\chi^2(1) = 4.75$, $p < .05$). Forty-one percent of the teachers also indicated to use the benchmark rating scale for writing instruction in class, which did not differ between trainers and trainees ($\chi^2(1) = 0.28$, $p = .60$).

During the interviews, trainers and trainees reported hardly any differences in how they implemented the key components of *Tekster*. They indicated that the writing strategy changed the content of their lessons. Because the lessons were divided into steps of the writing process, teachers focused more on the process of writing, instead of mainly on the product. Whereas the focus of instruction changed, it became clear during the interviews that teachers struggled with adapting their mode of instruction. Teachers who were used to apply modeling during instruction applied this also in their writing lessons. However, some of them indicated that they intentionally did not model, as they were afraid that students would not be able to come up with information themselves. Teachers who were not used to apply modeling, indicated that they would like to have had more training and practice in teacher modeling, as they found it sometimes hard to decide what to model and how to do it. One of the trainers who was already experienced in modeling, indicated that the training offered enough support to implement modeling effectively in the writing lessons and to instruct a colleague, but that this might pose a bigger challenge for teachers who are not yet familiar with the underlying principles of modeling. Moreover, when teachers provided support to students, for instance by modeling or feedback, this was mostly offered during the prewriting and writing phase. During the post-writing phase students worked mainly independently without extra teacher support. According to the teachers, this was due to lack of time. This was also the main reason for not evaluating all students' final writing products or providing feedback on a regular basis.

3.3.3. Implementation of the teacher training

From the post-intervention questionnaire it became apparent that the collegial training sessions differed from the expert training sessions, both in terms of duration as well as in the content that was discussed during the sessions. The two expert sessions lasted 8 hours in total, whereas the collegial training session consisted of two or multiple sessions of, on average, 93 minutes in total. However, the duration of the collegial training differed largely between teachers (*SD* = 83), and ranged between 0 and 480 minutes. For most of the teachers (71%) the training sessions were one-on-one. There were only four trainees who attended a *Tekster*-lesson of their colleague. The topics that were discussed during the collegial training sessions were the goal and structure of the lesson program (84%), organisation of the lessons (84%), feedback (81%), rating text quality (78%), modeling (76%), benchmark rating scale (73%), student texts (46%), and specific problems (46%).

That the content of the professional development program substantially differed for trainers and trainees became also clear from interviews with teachers. Teachers confirmed that the collegial training sessions were relatively short and mainly dealt with the content of the intervention and organizational issues. Moreover, trainees did not study the provided background information in the teacher manual. As a result, trainees were less aware of the importance of the topics that were covered during the professional development program, such as modeling, feedback and assessment of text quality. This can partly explain why trainees struggled with changing their mode of instruction. It is quite remarkable that trainers did not transfer their newly acquired knowledge about effective instructional modes, as trainers reported that they found the information concerning modeling, feedback and assessment of text quality particularly useful. For instance, one trainer stated that benchmarks helped students to “understand what

constitutes a good text", which motivated students to revise their own texts. A possible explanation why trainers did not address the mode of instruction is perhaps that they overestimated their colleagues. For instance, one trainer stated that she did not put a lot of effort in the training session with her colleague, as "she is a very experienced teacher". The trainees themselves indicate that they hardly needed any support from their trained colleague, as "the teaching materials and the manual were very clear".

3.4. Social validity

The acceptability of and satisfaction with the intervention was established by combining logbook data, questionnaire and interview data on teachers' experiences with the lessons, the key components of the program, the manual and the training sessions.

3.4.1. Teachers' experiences with the lesson program

With regards to the lesson program, logbook data revealed that, on average, teachers estimated the level of difficulty of the lessons for their students with 3.03 (SD = 0.47) on a five-point scale ranging from 1 (easy) to 5 (hard). This indicated that the lessons were challenging, but not too difficult for the students. Further, teachers reported on the level of difficulty of executing the lessons, also on a five-point scale ranging from 1 (easy) to 5 (hard). This aspect was rated with an average score of 2.39 (SD = 0.50), indicating that teachers felt sufficiently equipped to teach the lessons. For these variables, there were no differences found between trainers and trainees ($F < 3.28$, $p > .07$). Teachers generally were positive about the teaching program. In the logbooks, trainers were slightly more positive than trainees (respectively M = 7.93, SD = 0.49 and M = 7.70 on a 10-point scale, SD = 0.52, $F(1,93) = 5.58$, $p = .02$). The post-intervention questionnaire showed that trainers and trainees were equally positive about the program

as a whole (M = 4.10 on a 5-point scale, SD = .62, $F(1, 62) = 3.14$, p = .08). Their students were also positive about the lessons (M = 6.75 on a 10-point scale, SD = 2.23).

The interview data confirm the quantitative data. Especially the communicative goals of the writing tasks during the lessons were frequently mentioned as a positive aspect of the program. In contrast to the writing tasks from the language textbooks that they are used to work with, *Tekster* offered them writing tasks with clear communicative goals that were close to the experience and interest of their students. Trainers as well as trainees indicated that these kind of writing tasks helped them to focus more at the content of students' writing instead of on formal aspects, like mechanics and conventions, during instruction. Moreover, because the writing tasks were close to students' own experience and the content of their writing was taken more seriously by the teacher, students seemed to be more motivated to write. However, this motivation was not always evident. Both trainers and trainees from different schools indicated that students sometimes started demotivated, "bleeeh, *Tekster* again!*"*. This changed to a more positive attitude during writing, especially when they were working together with peers.

3.4.2. Teachers' experiences with the key components

Teachers indicated in the questionnaire that the key components of the writing program such as the strategy, modeling, model texts, feedback, and peer interaction, supported their students' writing performance, see also Table 7. We found no differences between trainers and trainees. Moreover, both trainers and trainees indicated that they intended to keep using components of the intervention program, such as modeling (84%), strategy-instruction (84%), the acronym (81%), discussing model texts (70%), peer interaction (78%), feedback (92%), and specific writing tasks from the lesson program (70%). The majority of trainers did also indicate that they intended to continue using the

benchmark rating scales for evaluating text quality (89%), whereas for trainees this was only 50%. Peer modeling using video clips was the only component of which only a minority of the teachers (14%) intended to continue using it.

The interview data made clear that teachers were positive about all the key components of the program. According to them, the writing strategy was the most effective component. Both trainers and trainees indicated that students experienced much support from the steps of the writing strategy. For instance, according to one teacher the benefit of the program is that "students are taken through the process of writing a text, step by step". Another teacher compared the steps to "a recipe that always results in a delicious cake". Teachers indicate that especially struggling writers experienced a lot of support by the strategy, because "it helped them to commence with writing more easily". It also turned out that the overall strategy promoted transfer: students continued to use the steps for writing tasks also after the intervention period. The support that teachers experienced from the writing strategy might also explain the improvement in their self-efficacy for teaching writing. One teacher reported that "because you are so focused on the writing process of your students, you start reflecting on your own writing process as well".

Teachers were less positive about the videoclips in which peers modeled (parts of) the writing process or the writing product, and about the peer-feedback component that was included. They indicated that the quality of the videoclips was insufficient. Some teachers reported that they stopped using the videos, but incorporated this information in the plenary instruction or in a classroom discussion. With regards to peer-feedback, both trainers and trainees indicated that their students lack the skills and knowledge to provide effective feedback on each other's texts. They revealed that when students provide feedback, they focus mainly on positive aspects in a very general way, such as "well

done", or they provide suggestions which are related to lower order aspects of the text, such as "add a title", or to formal aspects such as "write more neatly". Both trainers and trainees indicated that students struggled with providing feedback on the content or communicative goal of the text.

3.4.3. Teachers experiences with the teacher manual and the training sessions

Table 7 presents the questionnaire data on teachers' satisfaction with the professional development activities. In general, teachers were highly satisfied with the teacher manual as well as with the training sessions, as they indicated that both supported them in the execution of the writing lessons. More specifically, the training sessions supported their practice in modeling, giving feedback and evaluating the quality of their students' texts. There was, however, a difference between teacher-trainers and teacher-trainees: trainers experienced generally more support from the expert training sessions than trainees did from the collegial training sessions (F (1, 62) > 6.47 and $p < .05$).

It is important that skills and knowledge on how to implement key features of a new program are easily transferable between teachers as in elementary education there often are changes in staffing, due to the large number of part-time working teachers and personal circumstances such as maternity leave or sick leave, which was confirmed by the interviewed teachers. In the interviews trainers indicated that it was easy to transfer the content of the training sessions to their colleagues and trainees reported that they experienced enough support from the trainers and the teacher manual to execute the *Tekster* lessons. Despite this, the interviewed teachers also indicated that it often takes a couple of years to implement a program like this in the most optimal way. One of the trainers stated that "the longer you work with it, the better it gets".

4. Discussion

In this study, we examined the effect of the writing intervention program *Tekster*, including professional development activities for teachers, on students' writing performance and on teachers' self-efficacy and attitudes for writing and the teaching of writing. *Tekster* is a comprehensive program for writing for the upper elementary grades, combining strategy-instruction, text instruction, and the teaching of self-regulation skills with observational learning, explicit instruction, and (guided) practice to address both the focus and mode of instruction. Professional development of teachers was promoted by means of a teacher manual including lessons plans and training on how to optimally implement key components of the intervention in the classroom. To investigate whether the content of such a professional development program could be transferred between teachers, we applied a teachers-training-teachers approach in which half of the participating teachers (trainers) were trained by the researchers, and these trainers subsequently trained one or more colleagues (trainees).

The quasi-experimental results showed that *Tekster* improved students' writing quality significantly (ES = 0.55). After the intervention, students wrote better texts than students who were engaged in their regular writing activities. As *Tekster* was tested on a large scale involving 1365 students and 68 teachers from 25 schools, and students' writing performance was assessed with nine writing tasks in three genres, this effect is generalizable over students, teachers, and tasks. Moreover, the effect of the program maintained over time: two months after the intervention, students still wrote qualitative better texts than they did before the intervention. Results further show that there were no differences between the writing performance of students of teacher-trainers and teacher-trainees, indicating that trainers and trainees implemented the intervention with equal

effectiveness. Results also show that both trainers and trainees became more positive and felt more efficacious about teaching writing after the intervention. Their general attitude toward writing was not changed by the intervention program.

As teachers had to implement the intervention program in their own classrooms, we examined whether they implemented the program and the collegial training sessions with fidelity. We also investigated the social validity of the intervention program and the teachers-training-teachers approach, as this provides valuable information about the feasibility of implementation in classroom practice. We used an explanatory sequential mixed methods approach in which we collected and analyzed quantitative data from post-intervention questionnaires, logbooks, and classroom observations, followed up by qualitative data from focus group interviews with a sample of one third of the participating teachers. This combination of different approaches provides a more complete understanding of how teachers implemented and experienced the writing program in their classroom than either method on its own.

The mixed methods data revealed that teachers implemented the key components of the intervention program and that there were no differences between trainers and trainees in how they implemented the intervention. Trainers and trainees reported that their focus and mode of instruction changed by working with the intervention. They used a more process-oriented approach for teaching writing in which they incorporated modeling. Furthermore, they supported students' self-regulation by explicitly strategy instruction and focusing on communicative goals of writing. They also reported to have adapted their instruction to the need of individual students by providing feedback, both on the written product and the writing process.

Concerning the social validity of the intervention, experiences of both trainers and trainees with the lesson program, the teacher manual, and the training were highly

positive. They indicated to especially appreciate the strategy-focused approach in the writing lessons and the writing tasks with explicit communicative goals. According to the teachers, the key components of the program were effective and they intended to keep using them.

4.1 Implications for improving writing education in upper elementary grades

The intervention program *Tekster* combines several instructional practices that have proven to be effective in earlier research into one comprehensive program for writing. Although we did not measure the effectiveness of individual components, the mixed methods data on teachers' implementation and experiences with the components of the program provide valuable clues on what aspects of the program may have been especially effective for improving students' writing and what aspects may have been less effective. To improve writing education in upper elementary grades, it is important to know what works and what not. Therefore, we will elaborate on these aspects in more detail.

4.1.1. Effectiveness of strategy-instruction

In line with previous research (cf. Graham, 2006; Graham et al., 2012; Koster et al., 2015) the results from the present study show that strategy-instruction is an effective instructional practice to enhance students' writing performance. The application of a writing strategy reduces the number of cognitive processes that are active at the same time, which reduces students' cognitive overload and ultimately leads to improved writing performance (Graham et al., 2012; Kellogg, 1988). The interviewed teachers in this study confirm that the steps of the writing strategy offered students support to manage their writing process more effectively. They indicated that by generating and

organizing ideas before writing, students came up with more ideas than before the intervention and that students wrote longer and better texts.

The strategy that was the core of every *Tekster* lesson also elicited a change in teachers' classroom practice: the emphasis of teachers' instruction shifted from a product to process approach to writing. Teachers provided more support to students during the writing process, they modeled parts of the writing process, and provided feedback during the process. Interview data also revealed that teachers felt more able to differentiate their instruction to meet the needs of weak as well as proficient writers. This is confirmed by their self-reported feelings of efficacy for teaching writing: both trainers and trainees indicated that they felt better equipped to teach strong and weak students to write. Although teachers paid more attention to the writing process in general, the results suggest that this was especially true for the prewriting phase, and less so for the post-writing phase. In interviews teachers indicated that students often had not finished their text during regular lesson time, which meant that the final steps of the strategy were finalized during independent work time, with considerably less support of the teacher. This indicates that the post-writing phase is not implemented in the most optimal way. Further research should investigate how the post-writing phase can be implemented more effectively.

4.1.2. Effectiveness of the communicative goals for writing and benchmark scales

The results from this study also emphasize the importance of explicating communicative goals for writing, as well as writing tasks that are close to students' own experiences. Teachers stated that because the lessons were meaningful for students they were willing to put effort in writing their text, even when they were not motivated to write at the start of the lesson. In previous research it was established that setting goals

for writing increases self-regulatory skills in writing (Zimmerman & Risemberg, 1997). The results from the present study seem to be in line with this. Teachers indicated that students had a better understanding of why they were writing a text and for whom, and were more able to monitor progress towards these goals, which might explain the improvements in text quality.

The mixed methods data also reveal that the explicitly stated writing goals supported teachers' instructional practice. Teachers reported to be more aware of the goal of writing, which facilitated providing feedback on students' writing products. They also experienced support by the benchmark rating scales when providing feedback. In the interviews, teachers indicated that the example texts in the rating scale, which reflect the range of students' writing performance for each writing genre, helped them to evaluate the communicative effectiveness of students' written texts. The data from questionnaires and interviews revealed that teachers used these scales not only for evaluating the quality of students' texts, and providing feedback, but also for classroom instruction. However, there were differences between trainers and trainees in both the implementation and satisfaction with the benchmark scales. In general, trainers were more satisfied with the benchmark scales and they used these scales more frequently than trainees did. This suggests that specific training on how to use these benchmark rating scales is essential for an optimal implementation in the classroom.

4.1.3. Effectiveness of peer interaction

During *Tekster* lessons, peer interaction is implemented at multiple occasions: students interact with each other before they start writing, e.g., to jointly generate ideas, but also after writing in order to evaluate texts written by their peers. The experiences of teachers with this component of the lesson program were mixed. On the one hand,

teachers indicated that interaction between peers enhanced students' motivation to write because students experience that their text is meant for a reader, and that their text is actually read. On the other hand, whereas it was the aim of peer interaction to raise awareness of the effect of their text on a reader, teachers indicated that students found it difficult to give their peers adequate feedback. According to the teachers, students generally responded positively to their peers, and when they did provide critical feedback this was mostly directed on lower order aspects in the text, such as bad handwriting, and errors in spelling or punctuation. Teachers stated that students need more support in order to provide effective feedback to their peers. This can, for instance, be done by including explicit, directive questions in the evaluation step of the writing strategy focusing on the information about text structures and criteria for a good text that were discussed during the introduction of each lesson. An important role for the teacher during this phase is to model to provide feedback and show what effective feedback looks like. Thus, it is essential that teachers not only guide and support the prewriting and writing phase, but also offer support during the post-writing phase.

Previous research showed that teachers themselves also struggle with providing effective feedback that is adjusted to the needs of the students (Bouwer, Koster, & Van den Bergh, 2016c). Although we focused on how to provide effective feedback during the teacher training sessions, it was beyond the scope of this study to determine whether teachers actually provided more effective feedback to students' writing. Further research is needed to investigate specifically whether teacher training enhances teachers' feedback on students' texts.

4.1.4. Effectiveness of observational learning

To effectively improve writing education, both the focus and mode of instruction should be addressed (Hillocks, 1984; Graham et al., 2012, Koster et al., 2015, Rijlaarsdam & Couzijn, 2000). The focus group interviews revealed that teachers struggled to adapt the mode of instruction, and that they especially encountered difficulties with the implementation of observational learning during writing instruction. Observational learning is an important component in the mode of instruction of *Tekster*, and it is implemented in two ways: through teachers who model parts of the writing process in class and through video clips in which peers model how they write texts in specific genres. Previous research indicated that through observing a model performing (part of) a writing task while thinking aloud, students gain insight into the writing process. Moreover, it provides students with the opportunity to direct their full attention to the learning task as the learning-to-write is separated from text production (Graham et al., 2005; Rijlaarsdam & Couzijn, 2000; Zimmerman & Risemberg, 1977).

With regards to teacher modeling, the results from this study showed that teachers implemented modeling in their writing, especially during the prewriting and writing phase. Modeling was seldom applied during the post-writing phase: students often had to finish this last step independently. Our results further indicate that some teachers struggled to implement modeling effectively in their instruction. Teachers who were already experienced in modeling stated during the interviews that effective modeling needs extensive practice and training, and that the training that was offered to them may not have been sufficient for teachers who have less experience in modeling. The interview data also revealed that some teachers did intentionally not model, as they were afraid that it would limit students' creativity. Some teachers indicated that their students literally copied the text that was modeled without coming up with their own ideas. In their opinion, students did not learn from this. However, research on observational

learning suggests that even when students copy ideas from the teacher they still learn from observing the writing process (Schunk, 2012; Zimmerman & Risemberg, 1997). This aspect of modeling should be investigated in more depth in further research.

The peer modeling videoclips were the aspect of the program that teachers were least satisfied with. In the interviews and in the logbooks teachers indicated that the quality of the sound of the videos was insufficient. Further, teachers indicated that students found it difficult to identify with the persons in the videos, which might have distracted students from learning from the content of the videos. Thus, our results indicate that peer modeling in this study was not implemented in the most optimal way. Future research should examine into more detail how peer modeling can be operationalized and organized in a way that it promotes students' learning.

4.2. Professional development of teachers

The aim of including professional development in the intervention program was twofold: to support teachers in implementing the intervention more effectively and with more fidelity, and to improve writing education on the long term, beyond the intervention (McKeown, Fitzpatrick, & Sandmel, 2014). To induce lasting improvement in the way writing is taught, the skills and knowledge of teachers have to be enhanced. Training teachers in applying effective writing practices increases their feelings of self-efficacy for teaching writing, which is positively related to their quality of instruction (De Smedt et al, 2016). Our findings indicate that teachers' feelings of self-efficacy were enhanced, suggesting that teachers themselves felt better equipped to teach writing after the intervention. The combination of enhanced self-efficacy for teaching writing with improved lesson material might have elicited a change in teachers' classroom practice. Mixed methods data from observations, interviews and questionnaires revealed that they

adapted both their focus and mode of instruction. Moreover, these changes seem to last beyond the intervention: even when teachers no longer had the *Tekster*-lessons at their disposal, their students' writing performance did not significantly decrease compared to their posttest performance.

Furthermore, teachers who are trained by experts and teachers who are trained by colleagues were equally capable of effectively implementing a comprehensive intervention program in their daily classroom practice. We found no differences between trainers and trainees in student outcomes, in teachers' perceived feelings of self-efficacy, attitudes toward writing instruction, classroom practice and experiences with the program. From these results we can conclude that a professional development program can be transferred between teachers within the same school. This is in agreement with prior research on spillover effects of professional development (Borko, 2004; Lieberman & Friedrich, 2007, Sun et al., 2013).

The only significant difference between trainers and trainees concerns the perceived support of the training sessions. In general, trainers reported to experience more support from their expert training sessions than trainees did from their collegial training sessions. That the teachers trained by experts were more positive than their colleagues about the key components of the program, such as modeling, feedback, and assessing text quality, is not surprising, as there appeared to have been differences in the content of the training sessions. In the expert training sessions the focus was primarily on the key components of the intervention, whereas the collegial training sessions primarily addressed organizational aspects of the intervention, such as when components had to be implemented during the lessons. However, it should be noted that the information on the content of the collegial training sessions was based on self-report data from the post-intervention questionnaire and focus-group interviews.

Despite the differences in the content of the training, the lesson observations suggest that there were no differences in how teacher-trainers and teacher-trainees implemented the key components during their lessons: both trainers and trainees applied modeling and provided feedback. This suggests that the professional development program contributed to change in teachers' instructional mode to an equal extent. However, during observations was only tallied whether these key components of the intervention were present, and not how they were executed. The quality of applying these key components might differ between trainers and trainees, because interviews revealed that during the collegial training sessions trainees received little support in how to change their mode of instruction. Trainers, on the other hand, reported that these aspects were eye-openers to them during the expert training sessions.

Why is it harder for teachers to adapt the mode of instruction than the focus of instruction? Whereas changing the focus of instruction primarily requires good teaching materials, changing the mode of instruction requires behavioral change, which is a process that takes time, practice and support. As Desimone (2009) indicated, to be effective professional development should be of sufficient duration, both in span of time and in number of hours spent on the activities. Unfortunately, research has not yet established a rule of thumb, but at least 20 hours of contact time spread over a semester is suggested (Desimone, 2009). It might well be that to have more impact on teachers' practice, the professional development activities in this study should have been longer in duration and more intensive. Concerning the transferability of the adaptation of the mode of instruction between colleagues should be noted that effective transfer requires that trainers understand the goals of the professional development program and know how these goals can be achieved (Borko, 2004). The fact that the trainers themselves still struggle with adapting the mode of instruction might be an explanation for the fact that

this aspect of the professional development program was not transferred. Future research should therefore especially focus on the role of professional development for the adaptation of the mode of instruction and investigate how professional development should be arranged to effectively address the mode of instruction and under what conditions this aspect is transferable between teachers.

A limitation of this study is that, due to the design of the study, we cannot be certain that the similarities between trainers and trainees can be attributed to a spillover effect, as we did not include a 'material-only' control group of teachers. It might well be possible that only the materials of the program (i.e., the lessons and the manual) have led to improvement in students' achievement. For instance, interview data revealed that teachers experienced much support from the lesson plans in the teacher manual for preparing their lessons. However, teachers indicated that the training supported them in the implementation of the program in their classroom practice, which was reflected in the increase of the scores of perceived self-efficacy.

Further, we only tested one form of professional development: a teachers-training-teachers approach with two training sessions with content specific for this intervention. The results of this study can therefore not be generalized and conclusions are limited to this specific intervention study. More research is needed to examine the spillover effect of professional development and the effectiveness of the teachers-training-teachers approach.

Finally, regarding the mixed methods approach, it should be noted that we analyzed the quantitative and qualitative data separately, and used the qualitative data to explain our qualitative findings, as we obtained qualitative data from only a subsample of participating teachers. This means that we do not know whether teachers who were more positive about the intervention in the focus-group interviews, and/or applied the key

components of the intervention more frequently according to the observations, were also more effective in their writing instruction, reflected by better writing scores of their students. In future research this relationship should be examined in greater detail.

5. Conclusion

In the present study we examined the effectiveness of a writing intervention, *Tekster,* on a large scale with nearly 70 teachers who implemented the intervention in their own classrooms. We included a professional development program to support teachers' skills and knowledge in working effectively with the program. To make this possible on such a large scale, we applied a teachers-training-teachers approach in which half of the teachers had to transfer their new skills and knowledge to their colleagues. It is shown that *Tekster* is an effective program for teaching writing and that additional professional development seems to support teachers to effectively improve their writing instruction. The present study also shows that a teachers-training-teachers approach is promising to implement professional development on a large scale. In this approach, the content of a professional development program is transferred between teachers within the same school. All in all, this study provides valuable clues how the gap between research and classroom practice can be bridged to improve writing education.

References

Ball, D. L., & Forzani, F. M. (2009). The work of teaching and the challenge for teacher education. *Journal of Teacher Education*, *60*, 497-511.

Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational researcher*, *33*(8), 3-15.

Bouwer, R., Béguin, A., Sanders, T., & Van den Bergh, H. (2015). Effect of genre on the generalizability of writing scores. *Language Testing, 32*, 83-100.

Bouwer, R., Koster, M., & Van den Bergh, H. (2016a). *Effects of a strategy-focused instructional program on the writing quality of upper elementary students.* Manuscript submitted for publication.

Bouwer, R., Koster, M., & Van den Bergh, H. (2016b). *Benchmark rating procedure, best of both worlds? Comparing procedures to rate text quality in a reliable and valid manner.* Manuscript submitted for publication.

Bouwer, R., Koster, M., & Van den Bergh, H. (2016c). *'Well done, but add a title!' Feedback practices of elementary teachers and the relationship with text quality*. Manuscript submitted for publication.

Braaksma, M. A. H. (2002). *Observational learning in argumentative writing* (Unpublished dissertation). University of Amsterdam.

Braaksma, M. A. H., Rijlaarsdam, G., Van den Bergh, H., & Van Hout-Wolters, B. H. A. M. (2010). Observational learning and its effect on the orchestration of writing processes. *Cognition and Instruction, 22*(1), 1-36. doi:10.1207/s1532690Xci2201_1

Brindle, M. (2013). *Examining relationships among teachers' preparation, efficacy, and writing practices* (Unpublished doctoral dissertation). Nashville, TN: Vanderbilt University.

Broekkamp, H., & van Hout-Wolters, B. (2007). The gap between educational research and practice: A literature review, symposium, and questionnaire. *Educational Research and Evaluation*, *13*(3), 203-220.

Couzijn, M., & Rijlaarsdam, G. (2004). Learning to write by reader observation and written feedback. In G. Rijlaarsdam, H. van den Bergh, & M. Couzijn (Eds.), *Effective teaching and learning of writing. Current trends in research* (pp. 224-252). Amsterdam: Amsterdam University Press.

Creswell, J. W., & Plano Clark, V. L. (2011). Designing and conducting mixed methods research (2nd ed.). Los Angeles, CA: SAGE Publications.

Department for Education. (2012). *What is the research evidence on writing?* (Research Report No. DFE-RR238). Retrieved from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/183399/DFE-RR238.pdf

Desimone, L. M. (2009). Improving Impact Studies of Teachers' Professional Development: Toward Better Conceptualizations and Measures. *Educational Researcher*, *38*(3), 181–199. doi:10.3102/0013189X08331140

De Smedt, F., Van Keer, H., & Merchie, E. (2016). Student, teacher and class-level correlates of Flemish late elementary school children's writing performance. *Reading and Writing*. Advance online publication. doi:10.1007/s11145-015-9590-z

Expert Group Learning Trajectories (2009). Referentiekader taal en rekenen: De referentieniveaus [Reference framework language and arithmetic: The referential levels]. Retrieved from the Ministry of Education, Culture and Science: http://www.taalenrekenen.nl/downloads/referentiekader-taal-en-rekenen-referentieniveaus.pdf

Fayol, M. (1999). From on-line management problems to strategies in written composition. In M. Torrance & G. Jeffery (Eds), *The cognitive demands of writing: Processing capacity and working memory effect in text production* (pp. 13-23). Amsterdam: Amsterdam University Press.

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

Graham, S. (2006). Strategy instruction and the teaching of writing: A meta-analysis. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds), *Handbook of writing research* (pp. 187-207). New York: The Guilford Press.

Graham, S., Harris, K., Fink, B., & MacArthur, C. (2001). Teacher efficacy in writing: A construct validation with primary grade teachers. *Scientific Studies of Reading, 5*(2), 177-202. doi: 10.1207/S1532799Xssr0502_3

Graham, S., Harris, K., & 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,* 207-241. doi:10.1016/j.cedpsych.2004.08.001

Graham, S., MacArthur, C., & Schwartz, S. (1995). Effects of goal setting and procedural facilitation on the revising behavior and writing performance of students with writing and learning problems. *Journal of Educational Psychology, 87,* 230-240. doi:10.1037/0022-0663.87.2.230

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,* 879-896. doi:10.1037/a0029185

Guskey, T. R. (1994). Results-oriented professional development: In search of an optimal mix of effective practices. *Journal of Staff Development, 15,* 42-42.

Harris, K. R., Lane, K. L., Graham, S., Driscoll, S. A., Sandmel, K., Brindle, M., & Schatschneider, C. (2012). Practice-based professional development for self-regulated strategies development in writing: A randomized controlled study. *Journal of Teacher Education, 63,* 103-119.

Henkens, L. S. J. M. (2010). *Het onderwijs in het schrijven van teksten* [*Education in text writing*]. Utrecht: Inspectorate of Education.

Hillocks, G. (1984). What works in teaching composition: A meta-analysis of experimental treatment studies. *American Journal of Education,* 133-170. doi:10.1086.443789

Holliway, D. R., & McCutchen, D. (2004). Audience perspective in young writers' composing and revising. Reading as the reader. In G. Rijlaarsdam (Series Ed.), & L. Allal, P. Chanquoy, & P. Largy (Vol. Eds.), *Studies in Writing: Vol. 13. Revision: Cognitive and instructional processes* (pp. 105-121). Dordrecht: Kluwer Academic Publishers.

Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research, 1,* 112-133. doi:10.1177/1558689806298224

Kellogg, R. T. (1988). Attentional overload and writing performance: effects of rough draft and outline strategies. *Journal of Experimental Psychology: Learning, memory and cognition, 14,* 355-365.

Koster, M., Bouwer, R., & Van den Bergh, H. (2014a). *VOS: werkboek en docentenhandleiding voor groep 6* [*FOX: workbook and teacher manual for grade 4*]. Utrecht: Utrecht University.

Koster, M., Bouwer, R., & Van den Bergh, H. (2014b). *DODO: werkboek en docentenhandleiding voor groep 7* [*DODO: workbook and teacher manual for grade 5*]. Utrecht: Utrecht University.

Koster, M., Bouwer, R., & Van den Bergh, H. (2014c). *EKSTER: werkboek en docentenhandleiding voor groep 8* [*MAGPIE: workbook and teacher manual for grade 6*]. Utrecht: Utrecht University.

Koster, M., Bouwer, R., & Van den Bergh, H. (2016b). *A letter of advice to Like: Examining the relationship between metacognitive knowledge and writing performance.* Manuscript submitted for publication.

Koster, M., Tribushinina, E., De Jong, P., & Van den Bergh, H. (2015). Teaching children to write: A meta-analysis of writing intervention research. *Journal of Writing Research 7*, 249-274. doi:10.17239/jowr-2015.07.02.02

Kuhlemeier, H., Til, A. V., Hemker, B., de Klijn, W., & Feenstra, H. (2013). *Balans van de schrijfvaardigheid in het basis- en speciaal basisonderwijs 2* [Present state of writing competency in elementary and special education 2] (PPON Report No. 53). Arnhem: Cito.

Lieberman, A., & Friedrich, L. (2007). Teachers, writers, leaders. *Educational leadership*, *65*(1), 42-47.

Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., . . . Busick, M. D. (2012). *Translating the statistical representation of the effects of education interventions into more readily interpretable forms* (NCSER 2013-3000). Washington, DC: National Center for Special Education Research, Institute of Education Sciences, U.S. Department of Education.

Luiselli, J. K., & Reed, D. D. (2011). Social validity. In *Encyclopedia of Child Behavior and Development* (pp. 1406-1406). Springer US.

McKeown, D., Fitzpatrick, E., & Sandmel, K. (2014). SRSD in practice: Creating a professional development experience for teachers to meet the writing needs of students with EBD. *Behavioral Disorders, 40*(1), 15-25. doi:10.17988/0198-7429-40.1.15

Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research, 62*(3), 307-332. doi:10.3102/00346543062003307

Pearson, P. D., & Gallagher, G. (1983). The instruction of reading comprehension. *Contemporary Educational Psychology, 8*(3), 317-344.

Pullens, T. (2012). *Bij wijze van schrijven: Effecten van computerondersteund schrijven in het primair onderwijs.* [In a manner of writing: Effects of computer-supported writing in primary education] (Unpublished doctoral dissertation). Utrecht University.

Rietdijk, S., Van Weijen, D., Janssen, T., Van den Bergh, H., & Rijlaarsdam, G. (2015). *Teaching writing in primary education: Classroom practices, learning time, and teacher characteristics and their relationships.* Manuscript submitted for publication.

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 Academic Publishers.

Rijlaarsdam, G., Janssen, T., Rietdijk, S., & Van Weijen, D. (in press). Reporting design principles for effective instruction of writing: Intervention as constructs. In R. Fidalgo, K. Harris, & M. Braaksma (Eds.), *Design principles for teaching effective writing: Theoretical and empirical grounded principles*. Leiden: Brill Publishers.

Salahu-Din, D., Persky, H., & Miller, J. (2008). *The nation's report card: Writing 2007* (NCES 2008-468). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education

Schunk, D. H. (1987). Peer models and children's behavioral change. *Review of Educational Research, 57,* 149-174.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston, MA: Houghton Mifflin Company.

Solomon, S. R. & Sawilowsky, S. S. (2009). Impact of rank-based normalizing transformations on the accuracy of test scores. *Journal of Modern Applied Statistical Methods, 8,* 448-462.

Sun, M., Penuel, W. R., Frank, K. A., Gallagher, H. A., & Youngs, P. (2013). Shaping professional development to promote the diffusion of instructional expertise among teachers. *Educational Evaluation and Policy Analysis, 35,* 344-369. doi:10.3102/0162373713482763.

Tschannen-Moran, M., Hoy, A. W., & Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research, 68,* 202-248.

Van den Bergh, H., & Eiting, M. H. (1989). A method of estimating rater reliability. *Journal of Educational Measurement, 26,* 29–40.

Van der Leeuw, B. (2006). *Schrijftaken in de lerarenopleiding: een etnografie van onderwijsvernieuwing* [*Written assignments in teacher training: An ethnography of educational reform*] (Unpublished doctoral dissertation). Utrecht University.

Van der Leeuw, B., Pauw, I., Smits, M. & Van de Ven, P., (2010). Schrijven op de pabo; wat weten we uit onderzoek? [Writing in teaching training college; what does research tell us?]. *Levende Talen Tijdschrift, 11*(2), 30-39.

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry, 17*, 89-100.

Yoshikawa, H., Weisner, T. S., Kalil, A., & Way, N. (2008). Mixing qualitative and quantitative research in developmental science: Uses and methodological choices. *Developmental Psychology, 44*, 344-354.

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

Zimmerman, B. J., & Risemberg, R. (1997). Becoming a self-regulated writer: A social cognitive perspective. *Contemporary Educational Psychology, 22*, 73-101. doi:10.1006/ceps.1997.0919

Table 1

Student information per grade and training group

Grade	Trainer group			Trainee group		
	N students	% female	Mean age *(SD)*	*N* students	% female	Mean age *(SD)*
4	178	49	9.19 (0.49)	283	49	9.16 (0.49)
5	197	49	10.28 (0.51)	230	50	10.24 (0.55)
6	227	49	11.24 (0.56)	250	50	11.23 (0.52)
Total	602	49	10.32 (0.99)	763	50	10.16 (1.01)

Table 3

Characteristics of Schools Selected for Observations and Interviews

	Schools				
	1	2	3	4	5
School location	Urban	Suburban	Urban	Rural	Rural
Student population	L1	L1	L2	L1	L1 with language deficiencies
Classroom organization	Single-grade	Multi-grade	Single-grade	Single-grade	Multi-grade
Teachers' appointment	Fulltime	Fulltime	Fulltime	Part-time	Part-time

Table 4

Fit and comparison of nested models

Model	N_{pars}	-2 LogLikelihood	Comparison Models	ΔX^2	Δdf	*p*
1 basic null model	5	83570.16				
2 + training group	6	83568.42	2 vs 1	1.74	1	0.19
3 + measurement occasion	8	83468.55	3 vs 1	101.61	3	< .001
4 + training*measurement occasion	9	78903.82	4 vs 3	4564.73	2	< .001
5 + training*measurement occasion (differential effects of intervention)	10	78902.32	5 vs 4	1.50	1	.22
6 + grade	12	78870.53	6 vs 4	31.79	2	< .001

Table 5

Average writing scores and their variances for students in the trainer and trainee group

	Parameter	*SE*	*t*	*p*
Fixed part				
Trainer group	91.77	1.12	81.94	<.001
Δt2	+7.29	0.58	12.57	<.001
Δt3	+0.36	0.64	0.56	.29
Trainee group	91.83	1.07	85.82	<.001
Δt2	+0.31	0.60	0.52	.30
Δt3	+7.29	0.58	12.57	<.001
Random part	S^2	*SE*		
Classes	32.57	6.43		
Tasks	15.66	1.26		
Students	50.45	2.47		
Error	78.44	1.19		

Table 6

Means and standard deviations of teachers' self-efficacy and attitudes towards writing and teaching writing

	Before	After
	M (SD)	*M (SD)*
Attitudes towards writing	3.13 (0.80)	3.05 (0.88)
Attitudes towards writing instruction	3.70 (0.73)	3.93 (0.55)*
Efficacy for teaching successful writing	3.30 (0.69)	3.56 (0.62)*
Efficacy for supporting struggling writers	3.55 (0.50)	3.72 (0.46)*

Note. * indicate that means were significant different over time.

Table 7

Means and standard deviations of teachers' satisfaction with the intervention

	Trainers *M (SD)*	Trainees *M (SD)*	Total *M (SD)*
Tekster lesson program			
General lesson program	4.24 (0.58)	3.97 (0.63)	4.10 (0.62)
Strategy instruction	4.59 (0.57)	4.44 (0.75)	4.51 (0.67)
Teacher modeling	4.41 (0.57)	4.24 (0.74)	4.32 (0.67)
Peer modeling in video-clips	2.83 (0.89)	3.06 (1.13)	2.95 (1.02)
Discussing model texts	4.31 (0.66)	4.15 (0.74)	4.22 (0.71)
Peer interaction	3.76 (0.87)	3.74 (0.99)	3.75 (0.93)
Feedback	4.17 (0.62)	3.91 (0.67)	4.03 (0.65)
Teacher manual			
Providing lessons	4.24 (0.87)	4.03 (0.87)	4.13 (0.87)
Using benchmark rating scale	4.62 (0.68)*	4.00 (1.16)	4.29 (1.01)
Training sessions			
Providing lessons	4.21 (0.77)*	3.06 (1.23)	3.59 (1.19)
Modeling	3.66 (0.90)*	2.79 (0.95)	3.21 (1.01)
Providing feedback	4.17 (0.82)*	2.59 (0.93)	3.32 (1.18)
Evaluating text quality	4.31 (0.71)*	2.97 (1.11)	3.59 (1.16)

* Means between trainers and trainees were significantly different, with F (1, 62) > 6.47 and $p < .05$; for all the other components means between trainers and trainees were not significantly different, with F (1, 62) < 3.14 and $p > .08$.

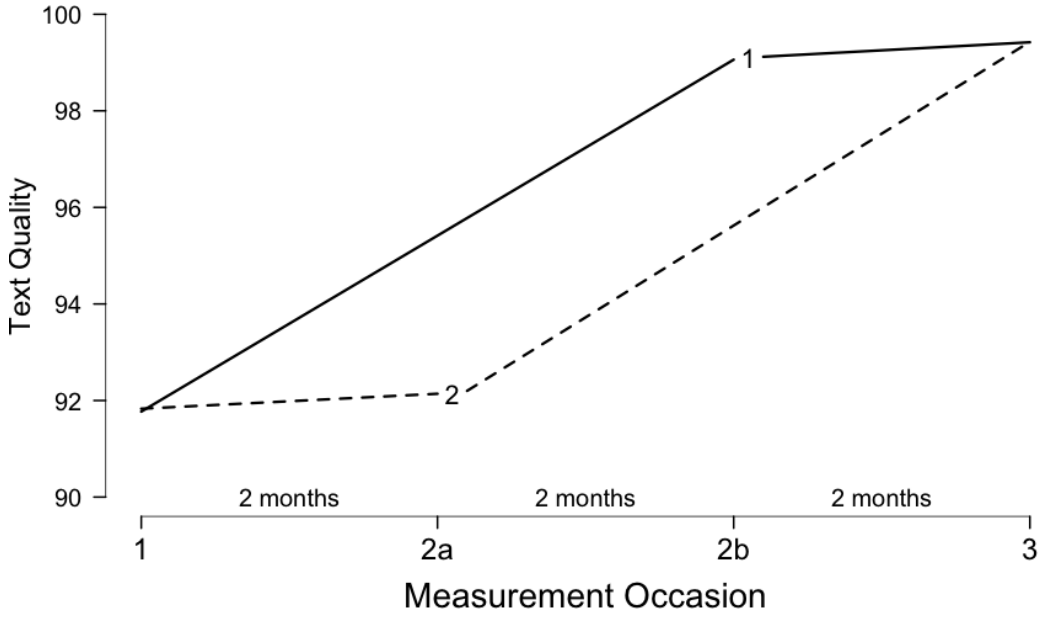

Figure 1. The effect of the intervention is estimated by comparing the slopes of the regression lines for text quality scores. Solid lines represent scores of the trainer group (1), who received the intervention between measurement occasion 1 and 2b. Dashed lines represent scores of the trainee group (2), who received the intervention between measurement occasion 2a and 3.

Appendix A

Sample lesson, translated from Dutch

SILENT BALL

Goal of the lesson: writing game rules

INTRODUCTION:

In the previous lesson you have learned how to write a recipe. A text like this, that teaches you how to do something we call an **instructive text**. An **instruction** describes the steps you have to take to make, cook or assemble something.

Game rules also are instructive texts. If you are used to playing games, you know that rules are very important.

In the video you are going to watch, two students discuss the content of game rules. They mention important aspects that have to be included.

Write down the 5 most important aspects:

1.....

.....

2.....

.....

3.....

.....

4.....

.....

5.....

.....

Now you are going to play an exciting game. This game is called 'Silent Ball'. You first will get a short explanation, and then you are going to play it.
Have fun!

Assignment

Silent Ball is a fun game, which you probably want to play again. But in a while you have probably forgotten the rules of the game. That is why it is handy to write down the rules, then you can consult them if you do not remember them. Try to write it down in such a way that someone who does not know the game can play the game without any problems..

HOW ARE YOU GOING TO DO IT?

To write the game rules you use the steps of EKSTER (MAGPIE):

1. **E**erst nadenken (think first)
2. **K**iezen en ordenen (choose and organize)
3. **S**chrijven (write)
4. **T**eruglezen (reread)
5. **E**valueren (evaluate)
6. **R**eviseren (revise)

STEP 1: E VAN EERST NADENKEN (THINK FIRST)

You are collaborating with a partner. First read the assignment again and remember the game you just played. What really has to be included in the game rules? Write all your ideas down in keywords.

.....

.....

.....

.....

.....

.....

.....

.....

STEP 2: K VAN KIEZEN EN ORDENEN (CHOOSE AND ORGANIZE)

Fill in the scheme below together with your partner. Use keywords.

Preparation
.....
.....
.....
.....
.....
Course of the game
.....
.....
.....
.....
.....
.....
.....
Ending of the game
.....
.....
.....
.....
.....
.....

STEP 3: S VAN SCHRIJVEN (WRITE)

You have thought about the rules of the game and the order of the rules. Now, write your rules down. Note that they must be clear for someone who is going to read them.

STEP 4: T VAN TERUGLEZEN (REREAD)

Read your game rules one more time. In the introduction you have written down five important aspects that have to be included in game rules. Did you include them in yours?

.....

.....

STEP 5: E VAN EVALUEREN (EVALUATE)

Exchange your game rules with another duo and read the tekst they have written.

Can you play the game with their rules?

☐ yes

☐ no

Write down tips to improve the game rules.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

STEP 6: R VAN REVISEREN (REVISE)

Read the tips you have been given and revise your game rules. Write your revised version below.

Appendix B

Exemplary writing prompts, translated from Dutch

Descriptive writing prompt: Lost keyring

While she was shopping in the supermarket, your mother has lost her keyring. She is very sad and she desperately wants it back. You want to help her by putting a notice on the notice board in the supermarket. Write a notice in which you ask whether anyone has found the keys. Describe what the keyring looks like, and where and when your mother lost it. Remember to mention your name and address to make sure that the finder can contact you.

Narrative writing prompt: Monkey

You see 3 pictures. They are the beginning of an exciting story about a monkey in the zoo. How will this story end? Make up the ending of the story. Write down the story from beginning to end, and also think of a good title for your story.

Persuasive writing prompt: Amusement park

You and your classmates want a make a daytrip to an amusement park. Your teacher does not find this a good idea. Still, you want to do everything you can to try to make this happen. Write a letter to your teacher in which you try to convince him or her with good arguments to go on a daytrip to an amusement park with the whole class. Clearly state in your letter to which amusement park you want to go.

Appendix C

Benchmark rating scale for persuasive letters