Improving Writing in Primary Schools through a Comprehensive Writing Program

Rietdijk, S.; Janssen, T.; van Weijen, D.; van den Bergh, H.; Rijlaarsdam, G.

Improving Writing in Primary Schools through a Comprehensive Writing Program

Saskia Rietdijk[1,2], Tanja Janssen[2], Daphne van Weijen[2], Huub van den Bergh[3] & Gert Rijlaarsdam[2,4,5]

[1] CED-Group, Rotterdam | The Netherlands

[2] University of Amsterdam | The Netherlands

[3] Utrecht University | The Netherlands

[4] University of Antwerp | Belgium

[5] Umeå University | Sweden

Abstract: This study examined the effects of an innovative comprehensive writing program in upper primary education on students' writing performance and on teachers´ classroom practices, beliefs and skills. The program focused on the communicative nature of writing, on writing as a process, and on explicit teaching of five genre-specific writing strategies. It was implemented by 43 teachers in their regular classrooms (Grades 4 to 6, N = 1052), with three conditions: (1) a writing program condition, (2) the same program complemented by professional development sessions and coaching, and (3) a control condition in which teachers taught their usual writing lessons.

Students' writing performance was measured three times with multiple writing tasks. Data on teachers' practices, beliefs and skills were collected through lesson observations, interviews, questionnaires, teacher logs, and a text assessment task.

The comprehensive writing program had a beneficial effect on students' writing performance and the extent to which teachers taught writing strategies. The complementary professional development and coaching had a direct effect on the number of lessons implemented, and an indirect effect on students' performance. Overall, the innovation proved to be effective for improving students' writing performance in the upper grades of primary schools.

Keywords: writing instruction, writing performance, teachers' beliefs, teachers' skills, primary school

1. Introduction

In the Netherlands, serious concerns have been raised about the level of students' writing proficiency at the end of primary school. Two national assessment studies indicated that two-thirds of the students in Grade 6 do not meet basic writing standards; their texts showed severe flaws with regards to content, organization, style, and communicative effectiveness (Krom, Van de Gein, Van der Hoeven, et al., 2004; Kuhlemeier, Van Til, Hemker, De Klijn, & Feenstra, 2013). Furthermore, students hardly progressed between Grades 4 and 6 (Kuhlemeier et al., 2013). These findings are alarming, since writing skill is crucial for students' academic success, and one of the basic requirements for participation in society.

Surveys via questionnaires and classroom observations indicate that little time is devoted to writing and writing instruction in Dutch primary schools, and that many teachers do not succeed in teaching writing effectively (Franssen & Aarnoutse, 2003; Kuhlemeier et al., 2013; for a review see Bonset & Hoogeveen, 2007). The Dutch Inspectorate of Education found that writing instruction was of insufficient quality in two-thirds of the schools (Henkens, 2010). Moreover, teachers are hardly trained to teach writing during their own professional development (Van der Leeuw, 2006). All in all, many students are struggling to write, and it seems that their teachers are often at a loss how to help them.

The present study aimed to contribute to the improvement of writing education in the upper grades of Dutch primary schools (Grades 4 to 6, age 9-12). For this purpose we designed, implemented, and tested a comprehensive writing program, including teacher materials and a complementary professional development (PD) program, accompanied by individual teacher coaching. The comprehensive writing program was implemented by teachers in their own classroom settings. The PD program and coaching aimed to support teachers to implement the writing program in the upper grades of primary school, to adapt it to their needs and local circumstances, and thereby to improve students' writing performance.

The study was set up as a research and dissemination study. That is, our aim was not only to test the effect of a particular writing program, but also to create a sustainable innovation. To this end, the writing program had to be feasible in regular practice, be practical for teachers to use, provide room for local adaptations, and be recognized as a program which matches teachers' curriculum expectations. It must be a practical operationalization of what teachers know the intended curriculum to be in this country. Therefore, we will also explain the status of the intended and realised writing curriculum as the context for the innovation.

Furthermore, giving teachers room to adapt the program is important, because research has shown that this can promote "the effective and sustained implementation of effective interventions" (Harn, Parisi, & Stoolmiller, 2013, p. 190), can help increase their motivation and engagement (Mowbray, Holter, Teague, & Bybee, 2003, p. 335), and that in some cases teacher modifications can have a positive influence on the

outcomes of an intervention (Dusenbury, Brannigan, Falco, & Hansen, 2003, p. 251-252).

In the next section we will outline the teaching of writing in the Netherlands, and the main principles on which the comprehensive writing and professional development programs are based.

2. Theoretical background

2.1 The teaching of writing in the Netherlands

We based the comprehensive writing program on three approaches to the teaching of writing: a focus on the *communicative* nature of writing, on writing as a *process*, and on the explicit teaching of *writing strategies*. Our choice was based on various sources: a survey of teachers' current beliefs, skills, and classroom practices in the upper grades of primary education (Rietdijk, Van Weijen, Janssen, Van den Bergh, & Rijlaarsdam, in press), consultations with experts in the field of writing education, and consultation of meta-analyses of writing intervention studies (e.g., Graham & Perin, 2007; Koster, Tribushinina, De Jong, & Van den Bergh, 2015).

The *communicative approach* to language teaching - sometimes called the 'functional' approach - has been advocated in the Netherlands since the 1970s (Ten Brinke, 1976; Leidse Werkgroep Moedertaaldidactiek, 1980). It arose as a result of democratization and other changes in society, and was developed in opposition to form-focused approaches to language teaching, which emphasize grammatical correctness and the use of 'fill in the blank' spelling, style, and sentence construction exercises (Kroon, 1985). Rather than focusing on promoting grammatical correctness, the communicative approach aims to foster students' 'communicative competence'. This notion, introduced by Hymes (1972), refers to the ability to communicate effectively in a range of social situations. This ability involves not only linguistic competence, but also knowledge of complex social rules: the knowledge of why, how, when, and to whom it is appropriate to use certain utterances. The communicative approach sees using language as a social act, as purpose-driven communication in a social context.

For the teaching of writing this means that attention must be paid to the functions of writing (to inform, to persuade, to entertain etc.) and to the development of communicative writing skills in various (semi-)real situations. The approach is largely based on 'learning by doing', that is; students learn to write by writing in a variety of communicative genres, while paying explicit attention to the rhetorical purpose and their readers' needs. The communicative approach emphasizes the authenticity and meaningfulness of writing activities; students write with real purposes for real readers, and receive feedback from their readers, not just from the teacher (Hoogeveen & Van Gelderen, 2013). Besides 'learning by doing', 'learning by observation' also fits well in the communicative approach: student writers observe readers while they try to make sense of their texts, and thus put the effectiveness of their texts 'to the test' (Rijlaarsdam

et al., 2008, 2009). Writers are then stimulated to 'read-as-the-reader', which is a strong incentive to set goals to revise and draft anew in order to improve the communicative effectiveness of their writing.

In their meta-analysis of writing interventions, Koster et al. (2015) found a large beneficial effect of goal setting on the quality of students' texts (*ES* = 2.03), and a moderate effect of peer assistance during parts of the writing process (*ES* = .59). Other studies indicate that the quality of students' texts may benefit from experiencing 'reading-as-the-reader' (Evers-Vermeul & Van den Bergh, 2009; Holliway & McCutchen, 2004; Rijlaarsdam et al., 2008, 2009).

Although communicative writing has received much support from curriculum designers and Dutch language teachers since the 1980s, it is only partially implemented in primary schools. Via teacher questionnaires, interviews, and observations of writing lessons of 61 teachers, Rietdijk et al. (in press) found that most teachers pay little attention to communicative aspects of writing in the upper grades of primary school. Writing assignments are often artificial, and students' texts are usually only read and evaluated by the teacher with a focus on spelling and punctuation. This might be due to the fact that communicative aspects remain underexposed in commercial textbooks and other teaching materials for primary school (Franssen & Aarnoutse, 2003). By incorporating communicative writing in the comprehensive writing program, we aimed to support teachers in implementing this approach in their writing lessons.

The communicative approach can be considered a 'social turn' in the teaching of writing. More or less at the same time, during the late 1970s, another shift took place: a shift in emphasis from the *product* to the *process* of writing, influenced by cognitive views on writing. Flower and Hayes (1981) designed their well-known writing process model in the early 1980s, and Hayes presented an updated version in 1996 (Hayes, 1996). In this model writing was viewed as a problem solving process in which ideas are actively constructed to meet communicative goals. It involved a variety of cognitive operations, such as planning, formulating and transcribing, and revising. The model was adopted early on in the Netherlands in secondary and higher education (Bochardt, 1984; Rijlaarsdam, 1986). The most important result in practice was the distinction between the draft and final version; the notion that exploring ideas in a draft version, and communicating ideas in the final version are distinctive activities in class.

In some secondary and higher education text books detailed guidance was provided for sub-processes, formulated in step-by-step procedures, which we would now refer to as strategies. These procedures focused on how to select and organize information from sources, for instance, and on which text pattern is best suited for which communicative act, for instance for promoting or defending a measure (Braet, 1979; Drop & De Vries, 1976). The instruction consisted of an explanation of the procedure or the steps, sometimes followed by a demonstration, and then practice.

The Dutch procedural approach for writing texts shares some features with the process approach as delineated by Graham and Perin (2007), in particular its focus on planning, drafting, and revising, and on goal- and audience-oriented writing. Other

instructional elements in Graham and Perin's definition (such as an emphasis on students' ownership of their writing, self-reflection, collaborative work, and individualized instruction), are less prominent in the Dutch curriculum proposals from the 1980s. Graham and Sandmel (2011) conducted a meta-analysis of 29 intervention studies which tested process writing instruction as defined by Graham and Perin. They found a moderate, beneficial effect of process writing on the quality of students' writing in general education classes, Grades 1 to 12 (ES = .34). This corresponds to the average effect size of .32 for the process writing approach on the quality of students' writing in Grades 4-12, reported by Graham and Perin (2007). However, the process writing approach only proved to be effective when professional development was also provided. All in all, process writing appears to be '(…) an effective, but not particularly powerful approach for teaching writing to students in general education classrooms' (Graham & Sandmel, 2011, p.404).

There are indications that process writing instruction – despite its benefits – is not fully implemented in Dutch primary education. According to the Dutch Inspectorate most teachers use pre-writing activities, but only a third of them asks students to revise their texts (Henkens, 2010). Classroom observations confirmed these findings (Franssen & Aarnoutse, 2003; Rietdijk et al., in press). Given the crucial role of the process in writing, and the need for students to become more aware of the activities involved in writing, we incorporated elements of the process writing approach in the comprehensive writing program; a focus on planning, drafting and revising, and on collaborative writing.

The third approach, *writing strategy instruction*, is relatively uncommon in Dutch primary education, as shown by Rietdijk et al. (in press) through teachers interviews and classroom observations. This is in contrast to reading strategies which are commonly taught in Dutch primary and secondary schools. Writing strategy instruction involves the explicit and systematic teaching of strategies for planning, drafting, revising, and editing texts. The level of explicit instruction exceeds what is usually provided in the process approach, where students mostly apply sub-skills without much instruction on how to do so (Graham & Sandmel, 2011).

Writing strategy instruction is a promising, well-investigated approach. In their meta-analyses of writing interventions Graham, McKeown, Kiuhara, and Harris, (2012) and Koster et al. (2015) reported large positive effects of strategy instruction on the quality of primary school students' writing (ES = 1.02 and .96, respectively). Therefore, we decided to incorporate writing strategy instruction in our comprehensive writing program.

The three approaches – communicative writing, process writing, and writing strategy instruction – are not mutually exclusive, but partly overlap. For instance, the process approach and writing strategy instruction both include planning, drafting, and revising activities, although the amount of explicit instruction provided may differ. Both approaches may also include writing for real audiences and receiving feedback from real readers, which are key components of the communicative approach as we defined

it. However, communicative writing tasks are not necessarily part of writing strategy instruction. We assume that communicative writing can be strengthened by paying attention to processes, which in turn may benefit from explicit and systematic instruction of strategies.

2.2 The comprehensive writing program in a nutshell

In the comprehensive writing program we focused on communicative aspects of writing, on writing as a process, and on the teaching of genre-specific writing strategies for five functions or genres; description, instruction, explanation, argumentation, and narration. Figure 1 visualizes how the three approaches were linked in the comprehensive writing program.

Figure 1. Three embedded approaches to the teaching of writing
in the comprehensive writing program.

First, students are confronted with a communicative situation, as well as with the additional purpose of writing and the appropriate genre (the outer ring of Figure 1). Next, students are stimulated to split the writing process into more manageable sub-processes or phases, and to engage collaboratively in planning, drafting, and revising texts within that particular genre (middle ring). At the centre of the program are the genre-specific writing strategies, which function as aids for planning, drafting, and revising. The comprehensive writing program is described in more detail below, in section 3.3.

2.3 Implementing a sustainable innovation

To increase the likelihood that teachers would continue to use the comprehensive writing program after our investigation, and to ensure the sustainability of the program, we embedded it in the reading comprehension program *Nieuwsbegrip* (*Comprehending the News*), a program that is widely disseminated, as 75% of the primary schools in the Netherlands has a license to use it (personal communication, educational center CED-Group, February 21, 2017). This online program provides a reading unit every week, in which students read a current news bulletin, for instance about the opening of the Dutch parliament, the Olympic Games, or Organ Donor Week. Students are taught reading strategies for comprehending these news bulletins. Subscribers can download the learning materials and manuals from the *Nieuwsbegrip* website (https://www.nieuwsbegrip.nl/). Teachers can print the material, or let students work with it on computers.

During the intervention, writing units were provided every other week, as an extension of the reading program. They were designed and produced by *Nieuwsbegrip*'s writing education experts, in close collaboration with members of the research team.

The advantages of embedding the comprehensive writing program in the digital *Nieuwsbegrip* environment were manifold. First of all, students wrote about current affairs topics, which may interest and motivate them. As the news item was introduced in a reading unit prior to the writing unit, students already had some knowledge about the topic they had to write about. Second, the teachers who participated in our study were already familiar with teaching reading strategies in the reading program, and had experience in modeling strategies, a key component of writing strategy instruction. Third, the writing program is flexible: although the design principles are fixed, the writing topics vary, depending on the news, and improvements can be made to the program along the way, based on users' feedback. Fourth, the format and lay-out of the lesson materials, and the means to distribute the materials (via the *Nieuwsbegrip* website) were already in place. And finally, perhaps most importantly from a viewpoint of sustainability, writing lessons continued to be designed and provided as part of *Nieuwsbegrip*, after the intervention ended and the research grant expired.

2.4 Professional development and coaching

All participating teachers were supported during the implementation of the comprehensive writing program with various activities and materials: there was an introductory meeting in which the principles and structure of the writing program were discussed, instructional materials (detailed lesson plans, writing tasks, feedback forms, etc.) were provided on the *Nieuwsbegrip* website, and teachers received a teacher manual with each new unit, which included modeling scripts when relevant.

However, providing materials might not be sufficient for an effective, sustained implementation of the new comprehensive writing program. Research indicates that professional development (PD) is crucial to school reform and improvement of

classroom practices (Garet, Porter, Desimone, Birman, & Yoon, 2001; McKeown, Brindle, Harris, et al., 2016; Opfer & Pedder, 2011). By participating in a high-quality professional development program, teachers can increase their knowledge and skills, and change their attitudes and beliefs. They can then use their newly acquired knowledge, skills, attitudes, and beliefs to improve their teaching, and thereby foster increased student learning (Desimone, 2009).

Studies of professional development indicate that successful, high-quality PD has the following five core features: (a) a focus on subject matter content, and on how students can learn that content; (b) opportunities for teachers' active learning: e.g. inquiry, discussion, planning, practice, observation of good practices; (c) coherence or alignment with teachers' own goals for learning and their goals for students; (d) co-operation or exchange between teachers; and (e) sufficient time for teachers to integrate their new knowledge into their classroom practices (Desimone, 2009; Garet et al., 2001; Penuel, Fishman, Yamaguchi, & Gallagher, 2007; Van Veen, Zwart, Meirink, & Verloop, 2010). In addition, individual coaching as follow-up to a professional development course has been recommended, to encourage accurate and sustained implementation of new teaching practices (Kretlow & Bartholomew, 2010).

The aim of our study was to develop and implement a professional development program based on these principles, to support teachers in implementing the comprehensive writing program, and to evaluate its effects on teachers' classroom practices, beliefs and skills, and their students' writing performance. Just as for the comprehensive writing program, the complementary professional development program continued to be provided by professionals of *Nieuwsbegrip* after the study expired, albeit in reduced form.

2.5 Teachers beliefs' and skills

Students' writing performance can be improved by changing teachers' classroom practices. Teachers' writing classroom practices in turn, are related to teachers' beliefs and skills. Research has shown, for instance, that there is a positive relation between teachers' efficacy and their use of evidence-based classroom practices (e.g., Brindle, Graham, Harris, & Hebert, 2016; Rietdijk et al., in press). Furthermore, a recent study indicated that teachers' ideas about writing instruction play a role as well: the more teachers valued correctness, the less they were inclined to give their students feedback on communicative aspects (Rietdijk et al., in press).

So, teachers' beliefs and skills may affect their choices in the classroom and how they implement new programs. The opposite is also true; practices can also change beliefs (Basturkmen, 2012; see also Borg, 2009). In other words, educational change is quite complex. Therefore, it seems prudent to examine the impact of the innovative writing program and complementary professional development program on teachers' practices, beliefs and skills in the domain of writing and writing instruction.

2.6 Research questions

In this study we examined the effects of the comprehensive writing program and the complementary professional development program on students' performance, classroom practices, and teachers' beliefs and skills in the domain of writing instruction. This enabled us to establish what changed as a result of implementing the programs. Furthermore, the outcomes might shed light on what contributed to the programs' effectiveness.

Our research questions were:

1. Is the comprehensive writing program more effective in improving students' writing performance than regular writing instruction?

2. Does the comprehensive writing program increase teachers' use of evidence-based practices and students' engagement in writing lessons, compared to regular writing instruction?

3. Does the comprehensive writing program have an impact on teachers' beliefs and skills regarding the teaching of writing, compared to regular writing instruction?

4. Do complementary professional development and coaching have additional beneficial effects on students' writing performance, teachers' classroom practices, and teachers' beliefs and skills in the domain of writing?

We expected that the comprehensive writing program would positively affect students' writing performance, teachers' classroom practices, beliefs and skills. For example, we predicted that teachers' use of specific approaches to the teaching of writing that are as yet not fully implemented in current practice (i.c. communicative writing, process writing, and strategy instruction) would increase. Furthermore, we expected that adding professional development and coaching to the writing program would have additional beneficial effects on students' writing performance, teachers' classroom practices, and their beliefs and skills in the domain of writing instruction.

The study took place during one full school year. It extends previous research on writing instruction by using a multicomponent approach, integrating communicative writing, a process approach, and strategy instruction, and by striving for a sustained innovation. In addition, we built on previous research by examining whether complementary professional development combined with individual coaching had added value for improving teachers' classroom practice and students' writing performance.

3.   Method

3.1   Design

The study involved an experimental design with three conditions: a comprehensive writing program only condition (WP), a writing program plus professional development condition (PD), and a control condition (C). In the WP and PD conditions teachers were asked to implement the comprehensive writing program in their own classrooms. In addition, the teachers in the PD condition participated in a professional development course and received coaching, to support them in implementing the writing program and help them improve the way they teach writing. In the control condition, teachers were asked to teach writing as they would normally do.

The measurement design consisted of three measurement occasions: a pretest, midtest, and posttest. An overview of all measures and measurement occasions is provided in section Measures.

3.2   Participants

Forty-three teachers and their students from 33 elementary schools participated in the study. With one exception, all teachers used the online reading comprehension program *Nieuwsbegrip*. In return for their participation, schools received a discount of 30% per teacher on their *Nieuwsbegrip* license. In addition, teachers received the results of their students on the writing tasks and a graph of students' on task behavior for each lesson observed.

The participating teachers taught Grade 4, 5 and/or 6. These grades correspond to student ages of 9 to 10 (Grade 4), 10 to 11 (Grade 5), and 11 to 12 years (Grade 6). About half of the teachers (47%) taught multi-grade classes, mostly a combination of Grades 5 and 6.[1]

Teachers were matched on the basis of gender, age, teaching experience, type of school, school location, and school quality according to the Dutch Education Inspectorate and grouped into trios. Then, each member of a trio was randomly assigned to one of three groups, which were subsequently assigned to conditions. At the start of the study six teachers were transferred at their own request to another condition, mostly because they found the PD condition too time consuming. As a result, there were fewer teachers in the PD condition ($N = 11$) than in the other two conditions ($N = 16$).

In total 1052 students of these 43 teachers participated in the study. At the beginning of the school year the children's parents or guardians received a passive consent letter, explaining the nature of the study. Parents of seven children indicated that they did not want their child to participate in the study.

We obtained background information (age, gender, parents' education level) for two thirds of the students, but data was missing for around a quarter of the students in the WP and control condition and 16% of the students in the PD condition. Students'

average age was 10 years (SD = 0.9), and half of the students (53%) were girls. Seventeen percent of the students' parents had a relatively low educational level, which means that both parents had completed no more than two years of high school.

Table 1 summarizes the participant demographics per condition.

There were no significant differences between conditions with respect to teachers' gender, χ^2(2, n = 43) = 1.07, p = .59, age, teaching experience or part time employment, F(6,78) = .17, p = .99. Nor were there significant differences in teaching multi-grade classes, χ^2(2, n = 43) = 2.33, p = .31, or the percentage of teachers who taught Grade 4 (only Grade 4 or in combination with other grades), χ^2 (2, n = 43) = 2.05, p = .36, Grade 5, χ^2(2, n = 43) = 5.96, p = .05, or Grade 6, χ^2 (2, n = 43) = 1.32, p = .52. School characteristics were similar across conditions: type of school, χ^2 (2, n = 43) = 5.27, p = .07, location of the school, χ^2 (2, n = 43) = .07, p = .97, or quality of the school according to the Inspectorate, F(2,35) = .47, p = .63, did not differ. However, the difference in school type and in teaching Grade 5 came close to being significant. Teachers who taught at public schools were under-represented in the PD condition. However, we have no reason to assume that teachers in public versus non-public schools teach writing differently.

Students' gender did not differ significantly between conditions, χ^2 (2, n = 800) = 1.48, p = .48, but students' age and grade level, F(4,1616) = 6.42, p = .00, and their parents' educational level did, χ^2(4, n = 811) = 29.39, p = .00. Fourth graders were over-represented in the PD condition, and their parents' education level was higher than in the other two conditions. Analyses also indicated a significant difference between conditions in students' reading comprehension, F(2,981) = 7.60, p = .00. However, this difference disappeared when we took students' age and grade level into account.

All in all, the three conditions were not equivalent in all respects at the start of the intervention. They differed in particular with regard to students' grade, and thus age. Therefore, we controlled for grade in the data analysis.

3.3 Comprehensive Writing Program

The intervention consisted of a comprehensive writing program (implemented in the WP and PD conditions), and a professional development program (PD condition only). At the start of the study an introductory session of three hours was organized for teachers in both intervention conditions, to acquaint them with the underlying principles, main components, and the structure of the comprehensive writing program.

The aim of the writing program was to improve the writing achievements of students in Grades 4 to 6 of Dutch primary schools. A unit of two consecutive writing lessons was designed every other week, by a team of educational designers from the educational center CED-Group in collaboration with members of the research team.

Table 1: Participant Demographic Information per Condition

Demographics	Writing program condition (WP) M (SD)	Professional development condition (PD)	Control condition
Teachers			
Number of participants	16	11	16
Male (%)	19	36	25
Age in years	43.2 (11.9)	41.8 (10.6)	42.2 (14.2)
Teaching experience in years	15.1 (9.7)	16.4 (10.0)	16.3 (13.9)
Employment (days a week)	3.8 (1.2)	3.8 (1.4)	3.8 (1.0)
Combination of grades (%)	56	27	50
Grade (single or in combination)			
Fourth (%)	56	64	38
Fifth (%)	56	46	88
Sixth (%)	38	18	25
Schools			
Number of schools	14	10	14
Type of school			
Public (%)	63	18	50
Non-public(%)	37	82	50
Location			
Urban (%)	50	55	50
Suburban / rural (%)	50	45	50
School quality *	3.3 (0.5)	3.1 (0.3)	3.3 (0.5)
Students			
Number of participants	381	272	399
Male (%)	48	44	49
Grade (%)			
Fourth	38	49	17
Fifth	33	40	73
Sixth	29	11	10
Age in years	10.2 (0.9)	9.9 (0.9)	10.3 (0.7)
Parents' educational level: 2 years of high school at most (%)	17	12	20
Reading comprehension**	27.6 (6.7)	29.8 (7.4)	29.2 (6.3)

*) School quality was rated by the Inspectorate on a four-point scale (1 = poor quality, 4 = good quality).

**) Reading comprehension was measured with a test developed by Aarnoutse and Kapinga (2006).

Each unit consisted of student materials (instructions, assignments), accompanied by a teacher manual. The teachers in the WP and PD conditions were free to incorporate the available materials at the moment and pace that suited them best.

In this section, we first describe the main components of the comprehensive writing program, followed by an outline of its structure and the specific teaching and learning activities that were incorporated in it. The PD program will be described in section 3.4.

3.3.1 Main components of the comprehensive writing program

As described in section 2, Theoretical background, the writing program integrated three approaches to the teaching of writing: communicative writing, process writing, and writing strategy instruction.

Communicative writing. We designed communicative writing tasks that students might encounter in real life, in and outside school. Texts were written for real purposes (as far as possible) and for real or semi-real audiences: classmates, friends, family members, or more distant others.

Five communicative purposes were taught during the year: to describe, to instruct, to explain, to persuade, and to entertain or express. Students learned to write texts in corresponding genres: descriptive, instructive, explanatory, argumentative, and narrative texts. These purposes and genres comply with the goals for the end of primary school, set by the Dutch Ministry of Education (Expert Group Learning Trajectories, 2009). Descriptive, explanatory and argumentative texts also play an important role in secondary education.

The communicative effectiveness of the text was the focal point of both instruction and feedback. The main focus was: does the text reach its communicative goal? To be able to improve the communicative effectiveness of their texts students were stimulated to ask their reader(s) for feedback, to observe readers trying to make sense of their texts ('testing your text'), and/or to observe readers evaluating their texts. Since previous research indicated that observation of readers is a highly effective learning activity within the communicative paradigm (Rijlaarsdam et al., 2008), we implemented it throughout the program.

Process writing. We stimulated students to engage in planning, drafting, and revising their texts individually, in pairs or small groups. These activities were linked to the communicative approach. In the planning stage, students oriented themselves on the topic, the rhetorical goal and the intended audience of the text. They generated ideas in various ways: by watching a television newscast about the topic, reading and talking about the topic, and/or brainstorming. In addition, they learned how to organize their ideas in the prewriting phase, for instance by ordering them in a list, a mind map or table. While drafting their first version, students engaged in collaborative writing with peers. Attention was paid to revising and rewriting, based on (peer) feedback from real

readers (see above: communicative writing). Students were also asked to reflect on the communicative effectiveness of their peers' texts and their own.

Writing strategy instruction. A central component of the comprehensive writing program was explicit and systematic instruction of strategies for planning, drafting, and revising texts. We chose not to teach general strategies (applicable to all kinds of texts), but opted for genre-specific strategies, which were applicable to one of the writing genres included in the study. For each genre a particular strategy was developed, to support students in coping with that genre (see Appendix A for an example of a genre-specific strategy). The ultimate aim was that students would be able to use these strategies independently.

Our writing program shared features with Harris and Graham's well-researched program of Self-Regulated Strategy Development (Harris & Graham, 2009). As in their model, each strategy was taught in six stages: (1) activating prior knowledge about writing in this genre; (2) direct, explicit instruction: explanation of the various steps of the strategy; (3) supporting memorization and retention of the strategy, through the use of mnemonics; (4) modeling of the strategy by the teacher; (5) practicing with support of materials, the teacher and/or peers; and (6) practicing independently, with little support. However, in contrast to the Harris and Graham model, *explicit* teaching of self-regulation skills was not part of our writing program. Instead, these skills were taught implicitly, through teacher modeling and by having students reflect on their writing process at the end of each unit (Fidalgo, Torrance, Rijlaarsdam, Van den Bergh, & Álvarez, 2015).

3.3.2 Structure of the comprehensive writing program.
The three approaches were integrated in each of the 21 units of the writing program. A unit consisted of two lessons of 45 minutes each in which students wrote one text. Three to five units were dedicated to each genre. At the beginning of the school year the units focused on descriptive and instructive texts. We assumed that these genres would be easier for students to master, and would contribute to audience awareness, important for communicative writing. Then the program moved on to explanatory and argumentative texts later in the year. For variety's sake, the informative genres were interspersed with narrative or creative ones (stories and poems).

The news topics were a constraining factor, which influenced our choices of genres in each period. Some news topics lend themselves well to writing in a particular genre (for instance, writing a descriptive text about an exotic animal which has gone missing), while other topics do not. As a result, the sequence of genres within a period depended to some extent on the actual news bulletins and was thus not fixed.

Each genre was taught in at least three different units, in a fixed sequence: first, students *experienced* the genre (unit 1), then they *acquired new procedural knowledge* (unit 2), and finally they were stimulated to *apply* that knowledge in new writing tasks (units 3-5).

Experience. The aim of the first unit of each genre was for students to experience the genre and get a sense of what was required, before learning a genre-specific strategy. Therefore, students were asked to write something that resembled the target text without much instruction, and then observed how readers responded to their texts.

This phase was linked to the communicative approach: by observing readers, students were made aware of the communicative effect of their texts. For instance, after writing a description of an object, student-writers were asked to observe readers (peers or the teacher) attempting to identify the described object among similar objects. If the readers failed to identify the object, students would experience that their text was not effective enough to steer the reader's mind and actions in the right direction: the communicative goal had not been met. Through this experience a need for explicit instruction in the genre was created and students became motivated to learn about it.

Acquire procedural knowledge. In a subsequent unit students learned a genre-specific writing strategy. The goal was that students would be able to reproduce the strategy steps at the end of the unit. A typical 'acquiring' unit consisted of five stages: 1) activating prior knowledge about (writing in) the genre, 2) direct, explicit instruction of the strategy, 3) supporting memorization of the strategy by providing a mnemonic, 4) modeling of the strategy by the teacher, and 5) practicing with support of the teacher and/or peers.

To activate students' prior knowledge the teacher recalled the experience unit and what students had learned about writing in the genre. Next, the teacher demonstrated and explained the genre-specific strategy, and provided mnemonics: one or two letters for each step of the strategy. For instance, the acronym SLAK (snail) was used for argumentative texts: describe the **S**ituation, **L**et your opinion be heard, provide **A**rguments, and end with a **C**racker of a sentence.

These mnemonics served two purposes. First, they helped students remember which steps to take in the writing process, so that retrieval was less demanding when they were engaged in writing. Second, mnemonics also provided meta-language for referring to the strategies, both during modeling by the teacher, and while practicing.

Subsequently, the teacher demonstrated the use of the strategy by modeling a relevant part of the writing process. A script was provided in the teacher manual, to support teachers' modeling. It was emphasized, however, that the script was just an aid, and that the teacher should model a 'natural' writing process as much as possible. This might include: making mistakes, voicing doubts and self-reinforcements, and revising ideas during writing. In this way, self-monitoring skills were taught implicitly through modeling.

Apply knowledge. The experience and acquire units were followed by one to three units in which students practiced the genre-specific writing strategy. First they practiced with support, for instance by working collaboratively in pairs or small groups, with help

from the teacher. Then, in a subsequent unit, students practiced independently while scaffolds were gradually withdrawn.

In a typical practice unit, the teacher first introduced the topic, the writing task, and the communicative situation. To activate prior knowledge, the teacher asked students what kind of text was required, for whom, and which strategy would be needed. Students were then reminded of the various strategy steps, after which they performed the writing task. From time to time students were asked to revise their texts, after receiving feedback from their peers. Peer feedback was supported by providing checklists with questions for students to use.

Students were encouraged to *integrate* their new knowledge and skills into their everyday life (cf. Merrill, 2002). We did not conceptualize this as a separate phase, but incorporated integration in all phases. At the end of each unit students were stimulated to reflect on their writing process and what they had learned. In addition, students' texts were published (e.g., in the school newspaper or on the school's website), so that students could publicly demonstrate their writing skill.

Figure 2 presents an overview of the consecutive learning units in the comprehensive writing program. The left and right handed panel present the invariant learning activities, while the mid panel provides the focus in successive learning units from top to bottom.
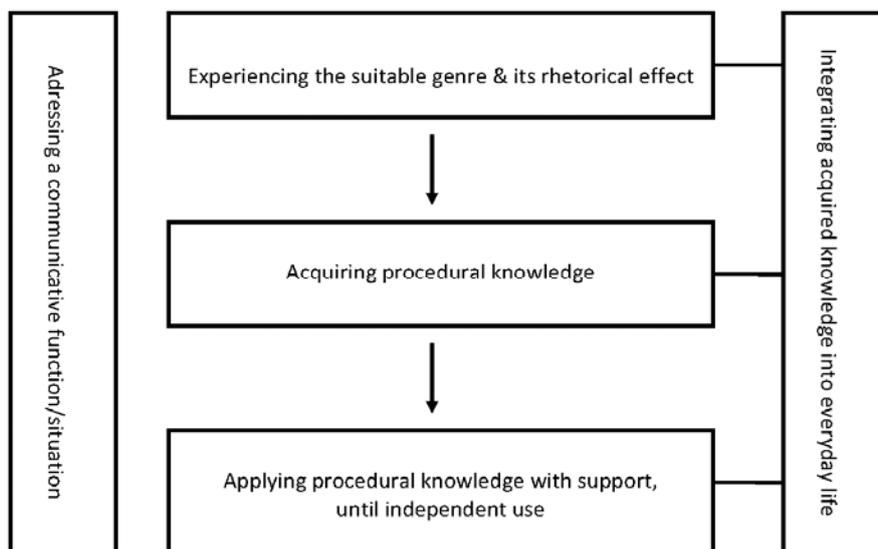

Figure 2. Main learning activities in the writing program.

3.3.3 Supporting materials

Various supporting materials were provided: classroom posters depicting each strategy, stickers on which students could fill in the strategy steps (so that they could look up the strategy if they needed it), and two folders for each student: one for collecting writing assignments and instructions, and one for their drafts and final written texts. Students also received checklists for providing peer feedback, with questions about communicative aspects of the text. Teachers received genre-specific rubrics in the teacher manual for assessing the communicative effectiveness of students' texts.

To enable teachers to differentiate according to the achievement level of their students, two versions of each unit were designed: an easier version intended for Grade 4 and for weaker writers in Grades 5 and 6 (level A), and a more difficult version for average and strong writers in Grades 5 and 6 (level B). In both versions the topic and genre were the same, but the units differed in complexity of the writing task, the number of strategy steps taught, the amount of instruction, and/or the level of support provided.

3.4 Professional development program

A professional development program was designed and implemented by two CED-Group writing education experts in collaboration with members of the research team. The aims of the program were to provide support during implementation of the newly developed writing program, and to stimulate teachers to engage in data-based inquiry and decision making in order to improve students' learning. This means that each participating teacher selected a problem or wish related to his or her teaching writing practice, and collected and analyzed data to help them make decisions on how to solve this problem and/or improve their writing lessons. The innovation was subsequently implemented and evaluated by the teacher.

The course consisted of six group sessions, each lasting four hours, and individual coaching in the workplace. Teachers in the PD condition participated in two small groups ($n = 6$ and $n = 5$) from September to June. They all began implementing the experimental writing program in their classrooms in August, at the beginning of the school year.

Between September 2013 and March 2014 five sessions were organized for each group, which was led by its own professional development leader, a writing education expert from CED-Group. These meetings focused on (a) the crucial elements of the writing program, and the choices teachers could make with respect to these elements, and (b) teacher knowledge and skills needed for data-based inquiry and decision making (e.g., how to assess the quality of students' texts, how to track students' writing development, and how to evaluate and adapt writing lessons within the program).

During the meetings the main principles underlying the writing program were discussed: communicative writing, process writing, writing strategy instruction, differentiating, and providing feedback. For instance, for *communicative writing*, a sample of student texts was assessed by the participating teachers, in small groups. This

led to a discussion of assessment criteria and the importance of focusing on the communicative effectiveness of students' writing.

With regard to *process writing*, the Flower and Hayes model of the writing process (Flower & Hayes, 1981) was presented and discussed, and linked to components of the writing program. In addition, different ways of implementing revision and collaboration activities (at different stages of writing) were exchanged.

How to implement *writing strategy instruction* was a central topic during the group meetings. A video of a teacher modeling a writing strategy was shown, and the teachers discussed several ways in which they could model writing strategies for their students through thinking aloud.

Teachers also practiced providing formative *feedback* on students' texts in small groups, and watched a video of students providing feedback on each other's texts. Characteristics of effective, helpful feedback were discussed, and the teachers collaboratively designed a checklist for providing written feedback.

The question how to *differentiate* one's writing instruction according to students' achievement level was addressed by discussing the difference between the two task levels (A and B), and by designing a so called group plan; an overview of the achievements of all students within a class, on different aspects of writing, and a plan which indicated how to adequately support and challenge the weaker and the better writers.

Between March and June 2014 the participating teachers carried out small scale inquiries in their own classroom, focused on a personal problem or wish with regard to their teaching of writing. In their inquiries teachers explored questions such as: 'How can I stimulate my students to revise their texts?' 'How can I adapt my lessons to better meet the needs of individual students?' and 'How can I provide helpful feedback during the writing process?'. While searching for solutions that fitted their context, teachers were coached by an experienced teacher trainer of CED-Group or a member of the research team. The coaches visited the teachers once or twice at their schools to observe writing lessons and to discuss teachers' questions and solutions. They also supported the teachers by e-mail, by asking questions, providing feedback, and offering suggestions.

During the sixth and final meeting the teachers presented the results of their classroom inquiries to the whole group. They also reflected on what they had learned about teaching writing from the training and coaching sessions, by writing a learner report and discussing their learning experiences in the group.

3.5 Measures

We used several instruments to measure the dependent variables: writing tasks, interviews, lesson observations, questionnaires, logs, and a text assessment task. Table 2 presents an overview of the instruments used per measurement occasion.

Table 2. Overview of Instruments used per Variable and Measurement Occasion

Variables	Pretest	Midtest	Posttest
Students' writing performance			
Writing tasks	x	x	x
Teachers' writing classroom practice			
Interview	x		x
Lesson observation	x	x	x
Questionnaire	x		x
Logs*			
Teachers' beliefs			
Questionnaires	x		x
Teachers' skills			
Logs*			
Text assessment task	x		x
Implementation			
Interview	x		x
Logs*			
Lesson observation	x	x	x

*Logs were filled in every two weeks throughout the year.

3.5.1 Students' writing performance

Students' writing performance was assessed on three occasions; before the writing program started, halfway through the program, and at the end of the program. We assessed students' writing performance with several tasks in different genres, because previous studies have shown that students' writing scores differ strongly across tasks (Bouwer, Béguin, Sanders, & Van den Bergh, 2014; Schoonen, 2012; Wesdorp, 1974). Students' writing was assessed in eight tasks in total (Table 3).

Table 3: Overview of Writing Genres per Measurement Occasion

| Genres | Measurement occasion | | |
	Pretest	Midtest	Posttest
Narrative	1		1
Expository	2		3
Descriptive	3	1	
Argumentative		2	2

Note: Numbers in the cells refer to the order of administration.

Narrative and expository writing tasks were used at both pre- and posttest (Table 3). We decided to administer the descriptive tasks at pretest and midtest, since this genre was taught during the first period of the writing program. Similarly, we administered the argumentative writing tasks at midtest and posttest, because this genre was taught during the second half of the writing program.

The tasks were adapted from previous studies (Bouwer & Koster, 2016; Krom et al., 2004; Pullens, 2012; Schoonen & De Glopper, 1996; Zwarts et al., 1990). The narrative writing prompts required students to write a short story about something scary (pretest) or a pleasant surprise they had experienced (posttest). In the expository task they were asked to write a letter of advice to a new classmate on how to write a good text (pre- and posttest). The descriptive writing task required students to describe a lost sweater for a lost-and-found department (pretest) or a cat that had gone missing for their neighbors (midtest). In the argumentative writing tasks students wrote a letter to persuade a company to grant them a present in the context of a promotion campaign (mid- and posttest).

The tasks were administered by trained assistants. A test session usually lasted about one hour. Students were not given a word-limit, but a time constraint was imposed: 15 minutes for writing a descriptive text (which was a short task), and 20 to 30 minutes for writing an expository, narrative or argumentative text.

All in all, 8025 student texts were collected and judged for their communicative effectiveness. For instance, would the neighbors be able to recognize the lost cat based on the student's description? Would the company be persuaded to send a present? To assess communicative effectiveness we used an approach developed by Blok (1986), in which texts are holistically rated by comparing them to five so-called anchor texts, which form a text quality scale. Anchor texts were selected from a sample of pre-rated texts. An average text was assigned the arbitrary score of 100. The scale contained two better and two weaker anchor texts, representing scores one or two standard deviations above and below the mean (scores 70-85-100-115-130). For expository and argumentative texts we used existing scales (Bouwer & Koster, 2016; Pullens, 2012). We used student texts collected by Pullens (2012) to create a scale for narrative texts, while for descriptive texts we created a scale based on a sample of texts from the present study.

Students' texts were split into randomly composed sets of about 50 texts per set, which were then randomly divided over 12 trained raters using a design of overlapping rater teams (Van den Bergh & Eiting, 1989). Each text was scored independently by three raters on communicative effectiveness, by comparing it to the anchor texts. The jury reliabilities (lambda) were satisfactory, ranging from .70 to .94.

3.5.2 Teachers' writing classroom practice

We collected data on various aspects of teachers' writing classroom practice. We expected use of the comprehensive writing program to result in writing lessons that included more elements of the three approaches: communicative, process and strategy

instruction. We also expected that participating in the writing program would lead to higher quality instruction, in terms of differentiation, strategy teaching, and promoting active learning.

These indicators were examined via stimulated recall interviews, lesson observations, questionnaires, and logs.

Interview. We designed an interview guideline, containing questions about (1) teachers' evaluation of the quality of students' texts, (2) the content and form of their writing lessons delivered in the context of language arts, and (3) the extent to which the teachers monitored their writing lessons.

Prior to the interview, the teachers were asked to send us copies of two student texts: a weak and a strong text. The interview started with a discussion of these texts, which focused on the teacher's criteria for assessing their quality (e.g. communicative aspects, style, structure, content), and the content and form of the lesson in which these texts were written. Most questions were open ended (for instance: 'Why is this text a weak text?', 'What happened in this lesson before students wrote their texts?'), followed by clarification questions (e.g., 'Did you provide instruction?', 'What kind of instruction did you provide?'). Follow-up questions were optional: whether they were asked depended on the teacher's response to the earlier questions. Finally, the teachers were asked whether they tracked students' writing development, and whether they designed and evaluated their writing lessons.

All 43 teachers were interviewed twice, at pretest and posttest, by trained research assistants (see 3.7 Data collection). On average, an interview took 41 minutes to complete (SD = 10). The interviews were audio taped and transcribed verbatim by the assistants.

The interview transcripts were subsequently coded to determine teachers' use of classroom practices: that is, whether or not they paid attention to communicative writing, process writing, and writing strategy instruction. Dichotomous codes were assigned: a teacher either reported using a practice (e.g., writing strategy instruction) or not. The text quality criteria mentioned by the teachers in the discussions of text samples were analyzed for attention to communicative aspects of writing, such as audience awareness and goal orientation, using a list of text criteria. A second coder rated the text quality answers in a sample of fifteen interviews. The inter-rater reliability was satisfactory (Cohen's kappa .88).

Lesson observations. Two writing lessons per teacher were observed by trained assistants at pretest, and one writing lesson was observed at mid- and posttest (one observer per lesson). In all, 171 writing lessons were observed and audio-taped. Posttest observation data for one teacher was missing because the observed lesson was not a writing lesson. Lessons were observed to determine students' time on task and teachers' classroom practices.

Time on task refers to the time spent by students engaged in learning tasks, as opposed to being off task (Karweit & Slavin, 1982). On task behavior is an indication of teaching quality, especially the extent to which the time for writing instruction is used efficiently. Students' time on task was measured by having a trained assistant observe eight randomly chosen students during each lesson ($N = 171$). Each student was observed for one minute, during which the observer scored twice (after 30 seconds) whether the student had been mainly on task or off task. A code for off task was assigned if the student was clearly not engaged in the lesson content (e.g. fidgeting or chatting). After all eight students had been observed for one minute, the assistant observed the same students again, in the same order. In a typical lesson eight students were each observed five times, resulting in 80 observation points per lesson (8 students x 5 minutes x 2 observations per minute). The percentage of students' time on task during the observed lessons was computed for each teacher per measurement occasion.

To measure teachers' *classroom practices*, we designed an observation instrument based on the Writing Observation Framework (Henk, Marinak, Moore & Mallette, 2003) and an observation instrument of the Dutch Inspectorate (Henkens, 2010). The instrument contained 25 questions, to be answered by the observer after the lesson had ended. Questions focused on whether or not the teacher used communicative writing, process writing, and writing strategy instruction. Examples of items are: 'Did the teacher provide a realistic writing task, situated in a real life context?', 'Did the teacher stimulate students to generate ideas before writing?', and 'Did the teacher pay attention to one or more writing strategies?'. Because the research assistants were occupied with scoring students' time on task during the lessons, we asked them to fill in the questionnaire afterwards, based on the audio recording of the lesson.

The reliability of the instrument was examined by having a second coder score a sample of ten lessons using the audio tapes. The inter-rater agreement between the observers and second rater was sufficient (Cohen's kappa .66 on average).

Questionnaire. An online questionnaire about providing high-quality instruction in general was administered to the teachers at pretest and posttest, containing three scales, based on Van de Grift (2007): Teaching learning strategies, Differentiating, and Promoting active learning. Van de Grift (2007) reported that these scales are positively associated with student engagement and achievement, and thus can be considered indicators of good and effective teaching.

The questionnaire contained 31 items, such as: 'Asking students to explain which strategy they use' (Teaching learning strategies), 'Adapting my writing lessons to students' different ability levels' (Differentiating), and 'Asking questions that encourage students to think' (Promoting active learning). Teachers were asked to indicate how often they engaged in these activities during their writing lessons on a five-point Likert scale (1 = never, 5 = always). The reliability of the scales was good (Cronbach's alpha's .80 - .87), see Appendix B.

Logs. To gather information on teachers' writing lessons throughout the year and on teachers' ability to reflect on lessons and adapt lessons to the context, we designed an online log. It contained questions about the writing lessons taught, the duration of these lessons, students' engagement, students' texts, teachers' satisfaction with and evaluation of the writing lessons (intervention conditions), and teachers' adaptation of the lessons (PD condition).

Teachers received an invitation via e-mail every fourteen days to fill in the log about the writing lessons they had taught in the previous two weeks. On average 14 logs were filled in per teacher (*SD* = 4.3).

3.5.3 Teachers' beliefs

We measured teachers' domain specific beliefs about writing and writing instruction, ranging from the views of what the act of writing means via their views about writing instruction, to their efficacy in teaching writing. We also measured their efficacy in aspects of high quality instruction, as indicators of general, non-domain specific efficacy. Data on beliefs was collected through questionnaires.

Teachers' *beliefs about writing* were measured with the Writing Beliefs Inventory (White & Bruning, 2005), which contains two scales: Writing as transmission (6 items) and Writing as transaction (13 items). Teachers with transmissional beliefs tend to view writing as a way to provide readers with information from authoritative sources. An example of an item is: 'The key to good writing is to report accurately on what experts think'. Teachers with transactional beliefs view writing as a way to personally construct the text by actively integrating their own thinking into the process. An item example: 'Writing helps me to understand the complexity of ideas' (White & Bruning, 2005). Teachers could respond to the items on a five-point Likert scale (1= totally disagree, 5= totally agree).

Teachers' *beliefs about writing instruction* were measured using the Writing Orientation Scale (Graham, Harris, MacArthur, & Fink, 2002), which contains three scales: Correct writing (5 items), Explicit instruction (4 items), and Natural learning (4 items). Teachers who score high on correct writing value formal correctness in students' writing. An example of an item is: 'Children should be reminded to use correct spelling.' Explicit instruction refers to the importance of direct skills-based instruction, e.g., 'It is important to teach children strategies for planning, checking, and correcting their texts.' Natural learning emphasizes the role of informal learning methods in the teaching of writing. For instance: 'Children gradually learn the requirements to which written texts should comply by writing and responding to others' texts'. Answers were given on a five-point Likert scale (1 = totally disagree, 5 = totally agree). We translated the questionnaire to Dutch and added two items per scale to increase reliability.

Teachers' sense of *efficacy in teaching writing* was measured with the Teacher Efficacy Scale for Writing (Gibson & Dembo, 1984; Graham, Harris, Fink, & MacArthur, 2001; Troia & Maddox, 2004). The Teacher Efficacy Scale is a 16-item

questionnaire, representing two dimensions: Personal and General teaching efficacy. Personal teaching efficacy refers to teachers' beliefs about their own ability to teach writing (e.g., 'When students' writing improves greatly, it is usually because I have found a more effective teaching approach'). General teaching efficacy refers to their beliefs about limitations on the effectiveness of teaching writing, created by environmental factors such as students' home environment (e.g., 'A teacher only has limited influence on students' writing performance; the students' home environment is more important'). The general teaching efficacy items were recoded, so that a higher score indicated a higher sense of efficacy, on a five-point Likert scale (1 = totally disagree, 5 = totally agree).

To measure teachers' *efficacy in providing high-quality instruction in general*, we designed a questionnaire based on an instrument developed by Van de Grift (2007) (see section 3.5.2). We used three scales from this instrument: Teaching learning strategies (7 items, e.g., 'Asking students to explain which writing strategy they use'), Differentiating (9 items, e.g., 'Adapting writing lessons to students' different ability levels'), and Promoting active learning (15 items, e.g., 'Asking questions that encourage students to think'). Teachers were asked to indicate how proficient they considered themselves to be at each activity on a five-point Likert scale (1 = not good at all, 5 = very good).

The belief questionnaires were distributed to teachers in an online environment. The response rate was 98% at pretest, and 95% at posttest. After deleting a number of items the reliability of the scales was fair to good, see Appendix B.

3.5.4 Teachers' skills

We measured teachers' skills in reflecting on their writing lessons, in adapting their lessons to the context, and in assessing the quality of students' texts. Teachers were stimulated to adapt the lesson materials to their local situation, and in the PD condition participants set up small scale design projects to adapt the materials to their specific needs. Therefore we included indicators of this kind of professional attitude and behavior in our instruments. Text assessment skill was included in the design, because being able to assess students' texts from a functional perspective could be a prerequisite for providing adequate feedback to students in the writing program we developed. The biweekly teacher manuals and the PD program focussed on how to provide feedback from a communicative perspective. Therefore, we expected that in the course of the school year, teachers might become more proficient in assessing texts from a functional instead of a formal perspective.

Teachers' *abilities to reflect on and to adapt lessons* to the context were measured through an analysis of the biweekly logs they filled in (see 3.5.2). Teachers were asked to indicate which aspects of their writing lesson(s) they were least and most satisfied with, and why. Two trained assistants coded the teachers' answers on three aspects: Perception (did the teacher perceive what went well or less well in the lesson), Analysis (did the teacher analyze the problem or source of success), and Adaptation (did the

teacher indicate that he or she had adapted the lesson). Teachers received a score for each item (1 = yes, 0 = no). Out of 342 completed logs 49 (14%) logs were coded by two independent coders to determine their level of agreement, which was high (Cohen's kappa .85).

Teachers' *ability to assess the quality of students' writing* was measured using a text assessment task. We asked teachers to rate 30 narrative and 30 argumentative texts, written by Grade 6 students. The teachers rated the texts holistically with the support of five anchor texts with fixed scores. The student texts to be rated and the anchor texts were obtained from Pullens (2012). Teachers were asked to focus on the communicative effectiveness of the texts. That is, how entertaining were the narratives, and how persuasive were the argumentative texts? The teachers rated the texts in a digital environment, at pretest and posttest; most teachers completed the task (pretest: 80%; posttest: 84%).

A jury of seven trained raters assessed the texts individually as well, in the same way as the teachers. The jury ratings were reliable (pretest: α = .90; posttest: α = .91). Subsequently, the teachers' ratings were compared to the average jury rating per text (correlations). We considered the correlation between the teacher's score and the jury's average score an indication of text assessment skill.

3.6 Implementation

In this study, fidelity measures have two functions. First, they indicate to what extent teachers implemented the writing program as it was intended, to validate the effects of the writing program found on students writing proficiency. Second, they can explain how differences in implementation of the program might influence the study's outcomes (O'Donnell, 2008; Kretlow & Bartholomew, 2010).

O'Donnell (2008) proposed five criteria for measuring fidelity: Duration (how many lessons were actually implemented?), Adherence (were intervention components delivered as designed?), Quality of delivery (how is the program delivered by the teachers?), Participant responsiveness (how are students and teachers engaged with the program's content?), and Program differentiation (are critical features of the intervention condition(s) present in those conditions and absent in the control condition during implementation?).

We measured all five aspects in one or more ways. With regards to Duration, we determined the number of writing lessons teachers taught. For the teachers in the control condition this was based on what they reported in the interview. The teachers in the intervention conditions were specifically asked which writing program lessons they taught, in the interview and the logs, and we used the highest number of lessons taught reported per teacher. For Adherence we determined the extent to which teachers in the WP and PD conditions adhered to key principles of the writing program during the observed lessons, e.g., whether or not teachers taught writing strategies and modeled the writing process when this was recommended in the lesson manual. For Quality of delivery, we determined whether teachers checked if their lesson achieved

its goal, which could be an indication that they cared about the effectiveness of their lessons. For Participant responsiveness we measured students' time on task behavior during the writing lessons. Finally, for Program differentiation, we examined whether features of writing strategy instruction were absent in the control group during the observed lessons.

Furthermore, to determine the success of the implementation of the professional development program we recorded teachers' attendance rate for the professional development meetings from meeting notes, and their overall satisfaction with the PD program and coaching sessions from their answers to interview questions.

3.7 Data collection

Data were collected within one academic year at three measurement occasions per teacher. Each class was visited twice on each occasion. Student and classroom practice data were collected between September and October (pretest), January and February (midtest), and May and June (posttest). The data collection for the teacher's profile started in May of the preceding school year.

The interviews and lesson observations were conducted by trained research assistants and the first author. They also administered the writing tasks. The assistants were university students. Most of them studied Dutch language, communication, psychology or educational sciences. They participated in a half-day training, during which they were informed about the aim of the study, read the interview guideline, viewed and discussed video segments of an interview, practiced interviewing each other, and received instructions for transcribing the interviews. They also practiced observing students' time on task by viewing, coding and discussing video segments of several lessons. Furthermore, assistants received a guideline for administering the writing tasks.

Teachers were interviewed during their lunch break or at the end of the school day. They were asked to authorize the interview transcript; no changes were proposed.

3.8 Data analyses

We collected data on the student level (writing performance) and the teacher level (classroom practices, beliefs and skills). The data have a complex hierarchical structure. Measurement occasions are nested within students, and students are nested within classes (teachers). As each student performed several writing tasks at each measurement occasion, writing scores are nested within writing tasks as well as within students. Hence, apart from variance due to error, there are three sources of variance: variance between writing tasks, students, and classes (see also Koster, Bouwer, & Van den Bergh, 2017).

Nine multilevel models were built to examine the effects of the comprehensive writing program and the professional development program on students' writing performance. In these models the hierarchical structure of the data, as outlined above, was taken into account. We started with a null model, which contained, next to the

four variance components, only an intercept (model 1). In model 2 we added the factor measurement occasion to test whether students' average writing score differed between the three occasions. In model 3 we included grade, followed by an interaction between grade and measurement occasion (model 4), allowing differences between measurement occasions to depend on grade.

In subsequent models we added the effect of the comprehensive writing program, which is used in both the WP and PD condition. In the first instance we tested whether the average writing scores differed between both experimental conditions on the one hand (WP and PD) and the control condition on the other hand, for the latter two measurement occasions (model 5). In this model the effect of the comprehensive writing program is assumed to be equal at the mid- and the posttest. This assumption was relaxed in model 6, in which the effect was allowed to differ between the mid- and the posttest. In model 7 we examined whether the effect of the comprehensive writing program differed between grades.

To investigate whether the professional development program had an additional effect, we allowed an interaction between the PD condition and measurement occasion in model 8. In the final model we tested whether the effect of the PD program differed between the midtest and the posttest (model 9).

The fit of the models is expressed by their -2 log likelihood. To test the significance of the distinguished effects, a likelihood ratio test was performed (Snijders & Bosker, 1999), as the difference in -2 loglikelihood between the nested models is χ^2-distributed with the difference in number of parameters as degrees of freedom.

Additionally, we tested the effect of the number of writing program lessons taught and the effect of the number of control lessons taught on students' writing performance, in a final model. Our expectation was that the more writing program lessons were taught, the higher students' writing performance would be, while at the same time the number of control lessons taught would hardly affect students' writing performance.

The data on the teacher level were analyzed through multilevel analysis as well. For each classroom practice, teacher belief and skill we built a multilevel model with measurement occasions nested within teachers in which we estimated the effects of the comprehensive writing program and the additional effect of the professional development program.

4. Results

4.1 Implementation

We did not ask the teachers in the intervention conditions to teach a certain number of writing program lessons. They were encouraged to teach as many lessons as they could, and were free to adapt the writing program to their own context and needs.

With regards to Duration, teachers in the intervention conditions taught on average eleven units of two lessons (*SD* = 3), so 22 lessons. This is about half of the units we provided. The teachers in the PD condition taught significantly more lessons on average

($M = 26$, $SD = 3$) than the teachers in the WP condition ($M = 20$, $SD = 7$), t (21) = 3.43, $p < .05$. The teachers in the control condition reported teaching 24 writing lessons ($SD = 14$) during the present study.

For Adherence, results indicated that 75% of the teachers in the intervention conditions were observed providing explicit instruction of a writing strategy at the midtest when this was recommended in the lesson manual, while 96% did so at the posttest. Seventy-three percent of the teachers in the intervention conditions modeled the writing process at the midtest during the observed lesson when advised to do so in the lesson manual. Modeling was usually part of the lessons in which a new strategy was introduced. As all strategies were introduced during the second unit per genre (see section 3.3.2), modeling was not recommended near the end of the year. Its implementation is therefore not reported for the posttest.

For Quality of delivery, we found that almost all teachers (91%) reported checking whether their lesson achieved its goal, while teachers in the intervention conditions did so significantly more often at the posttest than those in the control condition, F (2, 40) = 4,186, $p = .022$ (WP: $M = 1.00$, $SD = 0$; PD: $M = 1.00$, $SD = 0$; C: $M = .75$, $SD = .45$). In other words, they often took the time to check whether their lesson was effective, which suggests that they cared about the quality and effectiveness of the lessons they taught. Concerning Participant responsiveness we established that students were highly engaged in the comprehensive writing program: the percentage of students' time on task in the intervention conditions was 92% at the midtest and 88% at the posttest.

Regarding Program differentiation, the proportion of teachers in the intervention conditions who were observed teaching writing strategies was significantly greater than in the control group, F (2, 82) = 6,275, $p = .003$ (WP: $M = .59$, $SD = .50$; PD: $M = .68$, $SD = .48$; C: $M = .26$, $SD = .45$).

Finally, implementation of the PD program was satisfactory. On average teachers attended five out of six sessions (range: 3 - 6, $SD = .94$). They all attended the first session (kick-off). Teachers were asked to rate their satisfaction with the PD program and the individual coaching on a five-point scale (1: very dissatisfied, 5: very satisfied). On average, participants indicated that they were satisfied with both the PD program ($M = 4.6$, $SD = .5$) and the coaching ($M = 3.9$, $SD = .7$).

4.2 Effects on students' writing performance

We compared nine multilevel models to examine whether the comprehensive writing program (WP and PD condition) was more effective in improving students' writing performance than regular writing instruction, and whether the professional development program (PD condition) had an additional effect. To determine which model had the best fit, a likelihood ratio test was performed. The model fits and comparisons are shown in Table 4.

Table 4 shows that there was an effect of measurement occasion (model 2 versus model 1), indicating that students' writing performance differed between the pretest, midtest and posttest. Adding grade improved the model further (model 3 versus model

2): as we may expect, the scores differ from grade to grade. An interaction between grade and measurement occasion did not improve the model fit (model 4 versus model 3), which means that the differences between measurement occasions did not vary across grades.

Table 4: Fit of Multilevel Models of Students' Writing Performance. Factors: Measurement Occasion, Grade, Writing Program and Professional Development Program

Model	Nparameters	-2 LL	Comparison			
			Models	ΔX^2	Δdf	p
1: null	5	53470.26				
2: + MO	7	53440.87	2 vs 1	29.39	2	< .001
3: + Grade	9	53357.88	3 vs 2	82.99	2	< .001
4: + Grade*MO	13	53350.50	4 vs 3	7.38	4	.290
5: M3 + WP*MO (2&3)	10	53347.04	5 vs 3	10.84	1	< .001
6: + WP*MO (3)	11	53346.95	6 vs 5	.09	1	.760
7: M5 + WP*Grade	12	53343.70	7 vs 5	3.34	2	.190
8: M5 + PD* MO (2&3)	11	53343.83	8 vs 5	3.21	1	.070
9: + PD* MO (3)	12	53343.81	9 vs 5	3.23	2	.200

Note: MO = measurement occasion; M3 = model 3; M5 = model 5; WP = writing program; PD = professional development program

The incorporation of an interaction between use of the writing program (in the WP as well as in the PD condition) and measurement occasion resulted in a significant improvement to the best fitting model so far (model 5 versus model 3). This indicates an effect of the comprehensive writing program on students' writing performance compared with the regular writing instruction at the mid- as well as at the posttest. The effect did not vary between these two measurement occasions (model 6 versus model 5). Adding an interaction between the writing program and grade did not improve the model (model 7 versus model 5), which means that the effect of the program did not vary between Grades 4, 5 and 6.

No effect of the PD program was observed. Including an interaction between the PD condition and measurement occasion did not significantly improve the best fitting model so far (model 8 versus model 5), not even if we distinguish between the PD program's effects at the mid- and the posttest (model 9 versus model 5). Therefore, we will report the estimated combined outcomes of the intervention conditions, not of the WP condition and PD condition separately.

All in all, model 5 is the best fitting model. This model estimates students' writing performance per measurement occasion, for different grades, and tests the overall effect of the comprehensive writing program. In Table 5 we present the average writing scores and the variances estimated under model 5.

Table 5. Students' Average Writing Scores and Variances Estimated under Model 5 per Measurement Occasion (SE: Standard Error)

Factors	Pretest		Midtest		Posttest	
	β	SE	β	SE	β	SE
Fixed effects						
Control condition, Grade 5	91.29	.96	87.91	1.18	92.19	1.13
Δ Grade 4	-7.65	1.07	-7.65	1.07	-7.65	1.07
Δ Grade 6	6.75	1.18	6.75	1.18	6.75	1.18
Δ Writing program			4.10	1.22	4.10	1.22
Variance components						
S^2 (tasks)	29.39	3.14				
S^2 (students)	72.35	4.17				
S^2 (classes)	9.23	3.81				
S^2 (residual)	126.05	2.43				

The average writing score of fifth grade students in the control condition at the pretest equaled 91.29 (Table 5). The text quality of Grade 4 students was 7.65 points lower than the average score of Grade 5 students, whereas students in Grade 6 had a higher average writing score (+6.75) than fifth grade students. The effect of measurement occasion (Model 2) is located at the midtest – the average writing score of Grade 5 students in the control condition decreased to 87.91 –, probably due to the set of writing tasks used at measurement occasion 2. Since the interaction effect of grade and measurement occasion did not reach significance, the effect of grade does not vary throughout the year.

Most important is the effect of the comprehensive writing program, estimated as 4.10 (*SE* = 1.22). Hence, the average writing score of fifth grade students in the intervention conditions equals 92.01 (87.91 + 4.10) at the midtest and 96.29 (92.19 + 4.10) at the posttest. The effect of the writing program equals about half a grade effect: that is, six months.

The number of writing lessons taught varied considerably between teachers in the intervention conditions, but especially in the control condition (see section 4.1 Implementation). To test whether the effectiveness of the comprehensive writing program depended on the number of lessons taught and to examine the contribution of the number of lessons taught in the control condition, we added these variables as fixed effects to model 5. As a result, the model fit improved significantly, χ^2 = 8.61; *df* = 2; *p* = .01. Whereas no effect of the writing lessons on the quality of students' texts was found in the control condition, β =.02; *SE* = .14, each unit of the comprehensive writing program taught increased students' average writing score with .71 points (*SE* = .28), which is about one tenth of a grade effect. Thus, each additional lesson taught contributes to the effect of the comprehensive writing program.

Comparison of the estimated variance components of the null model and the final model indicates that 15% of the differences between tasks, 7% of the differences between students (within schools), and 69% of the differences between teachers are explained. So, at the teacher level a large part of the variance is explained, while at the student level nearly all variance remains unexplained. However, this is not surprising as the model does not contain any students' characteristics except for grade.

In conclusion, the comprehensive writing program was found to be more effective in improving students' writing performance than regular writing instruction. The more writing program lessons were taught, the more students' writing performance improved, in contrast to the control condition in which the number of lessons taught did not have an effect. However, an additional effect of the professional development program on students' writing performance could not be demonstrated.

4.3 Effects on teachers' classroom practices

We investigated the effects of the comprehensive writing program on teachers' domain-specific classroom practices (communicative writing, process writing, and writing strategy instruction) and on features of high-quality instruction (teaching learning strategies, differentiating, and promoting active learning). Furthermore, we examined the effect on students' time on task behavior, i.e. the extent to which students were engaged in instructional activities or learning tasks.

For each classroom practice we built a separate multilevel model, with observations nested within teachers. Just as in the models of students' writing performance it turned out that the PD program did not contribute significantly to the models. So, we report the results of the WP and PD conditions taken together. Table 6 presents the percentage of teachers who used elements of communicative writing, process writing and writing strategy instruction in their lessons at the pretest and the posttest.

Communicative writing. Less than half of the teachers mentioned the importance of goal directedness in the pretest interview (Table 6). At the end of the year the importance of goal directedness had increased substantially. Approximately one fifth of the teachers mentioned the importance of students' audience awareness. Classroom observations indicated that few teachers provided formative feedback on the goal directedness and audience awareness of students' texts. Approximately 80% of the teachers reported that students' texts were read aloud in the classroom. Furthermore, students' texts were published in some form in most classrooms.

Table 6: Percentage of Teachers who used Communicative Writing, Process Writing and Writing Strategy Instruction per Condition at Pretest and Posttest

Features of three approaches of teaching writing	Source	Intervention		Control	
		Pretest %	Posttest %	Pretest %	Posttest %
Communicative writing					
Goal directedness is mentioned while discussing text quality[a]	Interviews		89	25	75
Feedback on goal directedness	Observations	33	17	38	23
Audience awareness is mentioned while discussing text quality	Interviews	15	33	19	19
Feedback on audience awareness	Observations	37	17	38	15
Students' texts are read aloud	Interviews	89	89	87	63
Students' texts are published	Interviews	69	76	79	64
Process writing					
Generating ideas	Observations	85	70	87	73
Organizing ideas	Observations	63	70	56	47
Revising texts	Interviews	58	59	36	31
Writing strategy instruction					
Teaching writing strategies[b c]	Observations	30	85	50	20
Modeling the writing process	Observations	44	41	50	27

[a]: significant effect of measurement occasion

[b]: significant condition effect

[c]: significant interaction effect of condition x measurement occasion

No significant interaction effects (condition x measurement occasion) were found, which means that a significant effect of the comprehensive writing program on communicative writing practices could not be demonstrated (see Appendix C for statistics).

Process writing. A vast majority of the teachers encouraged students to generate ideas before writing (Table 6). Most of the teachers encouraged students to organize their ideas, as a pre-writing activity. Revising texts by students was implemented by a small majority of the teachers in the intervention conditions and by one third of the teachers in the control condition. We found no significant effect of the comprehensive writing program on process writing activities (see Appendix C).

Writing strategy instruction. According to the observations, the percentage of teachers who taught writing strategies in the intervention conditions tripled at the posttest, whereas the control condition showed a decrease (Table 6). This interaction effect was significant, F (1, 81) = 14.05; p < .001. Even when controlled for the multiple tests carried out (Bonferroni) the interaction between condition and measurement occasion is significant (against a critical p-value of .005). A minority of the teachers was observed modeling the writing process. No significant differences were found for modeling.

High-quality instruction. The means and standard deviations for three features of high-quality instruction are presented in Table 7.

On average teachers indicated at the pretest that they sometimes taught learning strategies. The mean score increased to quite often at the posttest, with the largest increase in the intervention conditions. However, the interaction effect between condition and measurement occasion was not statistically significant. Furthermore, teachers reported that they differentiated sometimes. Lastly, teachers indicated that they promoted active learning quite often. The mean score significantly increased at the posttest.

No significant interaction effects between the intervention conditions and measurement occasion were found for high-quality instruction (see Appendix C). So a significant effect of the comprehensive writing program on these features could not be demonstrated.

Table 7: Aspects of High-Quality Instruction per Condition at Pretest and Posttest (1: never; 5: always)

Types of High Quality Instruction	Intervention				Control			
	Pretest		Posttest		Pretest		Posttest	
	M	SD	M	SD	M	SD	M	SD
Teaching learning strategies [a][b]	3.39	.65	3.82	.38	3.18	.65	3.41	.71
Differentiating	3.03	.59	3.21	.56	3.14	.84	3.10	.77
Promoting active learning [a][b]	3.80	.46	4.03	.29	3.62	.41	3.72	.41

[a]: significant effect of measurement occasion

[b]: significant condition effect

Engagement. In Table 8 we present the means and standard deviations for students' time on task. The percentage of students' time on task was high throughout the year: around ninety percent. The interaction between condition and measurement occasion was not significant, $F_{(2, 80)} = 3.04$; $p = .05$, indicating that the intervention did not affect the level of students' engagement during writing lessons.

To summarize the effects on classroom practices, the comprehensive writing program had a significant beneficial effect on the extent to which teachers taught writing strategies. However, no significant effects of the writing program were found for communicative writing, process writing, teaching learning strategies, differentiating, promoting active learning or the level of students' engagement.

Table 8: Students' Time on Task per Condition per Measurement Occasion

Measurement Occasions	Intervention		Control	
	M %	SD	M %	SD
Pretest	89	10	85	12
Midtest	92	7	94	5
Posttest	88	9	94	6

4.4 Effects on teachers' beliefs

We investigated the effects of the comprehensive writing program on ten teacher beliefs (see Table 9). We built a separate multilevel model for each teacher belief. As the PD program did not contribute significantly to the models, the results of the intervention conditions as a whole are reported.

Table 9: Teachers' Beliefs per Condition at Pretest and Posttest

	Intervention				Control			
	Pretest		Posttest		Pretest		Posttest	
	M	SD	M	SD	M	SD	M	SD
Writing beliefs								
Writing as transmission	2.30	.50	2.15	.52	2.19	.57	2.27	.73
Writing as transaction	3.77	.33	3.76	.27	3.59	.52	3.67	.52
Writing instruction beliefs								
Correct writing	2.90	.60	2.73	.57	2.91	.53	3.02	.51
Explicit instruction	4.18	.35	4.19	.39	4.16	.38	4.28	.45
Natural learning	4.19	.41	4.42	.42	4.13	.36	4.13	.33
Efficacy beliefs								
Personal teaching efficacy	3.43	.39	3.60	.39	3.47	.39	3.51	.46
General teaching efficacy	3.39	.45	3.34	.52	3.52	.66	3.55	.48
Efficacy in teaching learning strategies[abc]	3.01	.71	3.70	.52	2.83	.65	3.10	.79
Efficacy in differentiating	2.81	.62	3.11	.58	2.86	.80	2.89	.87
Efficacy in promoting active learning[a]	3.49	.52	3.89	.26	3.41	.62	3.66	.60

Writing (instruction) beliefs scales: 1: totally disagree, 5: totally agree
Personal and General teaching efficacy: 1: totally disagree, 5: totally agree
Efficacy scales: 1: not good at all, 5: very good
[a]: significant effect of measurement occasion
[b]: significant condition effect
[c]: significant interaction effect of condition x measurement occasion

Table 9 shows the means and standard deviations for the teachers' beliefs scales. On average teachers held low transmissional beliefs and high transactional beliefs. Teachers neither agreed, nor disagreed on the importance of correct writing, but they valued explicit instruction and natural learning. They felt quite efficacious about teaching writing and this increased by the end of the year. Teachers felt most

efficacious about promoting active learning, and their confidence in this skill had increased by the end of the year. The average scores for efficacy in differentiating were below 3, which indicates that teachers have some doubts about their ability to differentiate during writing lessons.

For efficacy in teaching learning strategies a significant interaction between condition and measurement occasion was found, F (1, 42) = 4.45; p = .04. The teachers in the intervention condition felt more efficacious in teaching learning strategies at the posttest than the teachers in the control condition. Hence, the writing program had a significant beneficial effect on teachers' efficacy in teaching learning strategies. No significant interaction effects between condition and measurement occasion could be shown for the writing beliefs or writing instruction beliefs (see Appendix C).

In sum, the comprehensive writing program had a significant positive effect on teachers' efficacy in teaching learning strategies, but not on other efficacy beliefs to teach writing nor on writing beliefs or writing instruction beliefs. However, we must be careful as multiple comparisons were carried out. When correcting for the multiple comparisons (Bonferroni) the effect on teachers' efficacy in teaching learning strategies is not statistically significant.

4.5 Effects on teachers' skills

Table 10: Teachers' Skills per Condition per Measurement Occasion

Intervention						
Teacher skills	Pretest		Midtest		Posttest	
	M	SD	M	SD	M	SD
Text assessment	.65	.11			.65	.15
Reflection on lessons	.80	.21	.81	.20	.95	.14
Adaptation of lessons to context	.10	.19	.11	.18	.10	.31
Control						
Text assessment	.61	.10			.68	.11
Reflection on lessons	.79	.22	.82	.21	.80	.26
Adaptation of lessons to context	.08	.19	.01	.05	.08	.29

Text assessment scale: 0: no correlation with the jury - 1: perfect correlation;

Reflection on lessons: 0: never - 1: always

Adaptation of lessons to context: 0: never - 1: always

A separate multilevel model was built for each skill. Because the PD program did not contribute significantly to the models, the results of the intervention conditions are

reported together. The means and standard deviations of teachers' skills are shown in Table 10.

Teachers' text assessment skill correlated about .65 with the jury, on average, and was similar across conditions and measurement occasions. Across conditions, on average teachers often reflected on their writing lessons in the sense of observing and analysing, but seldom adapted their lessons to the context.

We found no significant interaction effects between the professional development condition and measurement occasion for teachers' skills. So, a significant effect of the professional development program on teachers' skills could not be demonstrated.

5. Discussion

In this study we examined the effects of a comprehensive writing program and a complementary professional development program on students' performance, teachers' classroom practices, and teachers' beliefs and skills in the domain of writing instruction, in the upper grades of Dutch primary education. We designed and tested a sustainable comprehensive writing program which intended to improve writing instruction and students' writing performance by combining pre-existing communicative and process oriented approaches to strategy instruction, an approach that proved to be effective in previous writing research (e.g., Koster et al., 2015; Graham & Perin, 2007; Graham et al., 2012). As this study did not only aim to improve students' writing performance through an innovative program, but also aimed to change teachers' instructional practices in the long term, the study took place in regular classroom settings. Participating teachers were free to incorporate the available materials at the pace and intensity that suited them best and to adapt the materials to their needs and local context. Implementation measures showed that on average teachers taught 22 lessons, about half the lessons the comprehensive writing program provided. Most teachers adhered to key principles of the program.

The comprehensive writing program was more effective in improving students' writing performance than regular writing instruction (ES = .27), and the number of lessons taught influenced the effectiveness of the program. This effect size is similar to the effect size found by Bouwer et al. (2016) in a study under similar conditions, in the same region (ES = .32). Meta-analyses of writing intervention research, however, showed much larger average effect sizes for writing strategy instruction in primary education: 1.02 (Graham et al., 2012) and .96 (Koster et al., 2015). These differences in effect sizes point to two issues related to calculating effect sizes.

First, there is the issue of measurement of writing performance. In line with Bouwer et al. (2016) ,we administered several writing tasks per student per measurement occasion, to estimate the writing skill on an individual level more precisely. The variance related to writing tasks is then part of the explained variance, which implies that the error component becomes smaller. However, most writing intervention studies measured students' writing performance with just one task. Because we modelled the different variance components, we know what the contribution of writing tasks to the

explanation of variance is. Therefore we simulate what the effect size would have been when we would have measured students' writing performance with only one task per occasion: if we ignore the variance caused by tasks the effect size would be doubled. In other words, this measurement issue affects the effect size because of a more precise estimate of the error components. Comparisons between studies in terms of effect sizes without taking into account the measurement issue is therefore problematic.

A second issue is that comparisons of effect sizes do not take into account what the research aims and conditions were. In the case of writing strategy instruction, most of the intervention studies included in meta-analyses were labeled as Self-Regulated Strategy Development studies (SRSD). SRSD is criterion-based, which means that instruction ends only if the student can use the strategy properly (Harris & Graham, 2009). However, the present study is restricted by the classroom schedule. Teachers taught the writing strategies in their regular classrooms, and were free to teach the number of lessons that best fitted their classroom practice. This is in line with what we aimed to achieve: to improve writing performance and writing practices in a naturalistic setting. As a consequence, the variation in implementation is larger than in controlled studies and the effect size is smaller, but the outcomes represent what can be attained in regular classrooms.

An important aim of the intervention was to change teachers' classroom practices. The writing program integrated three approaches: communicative writing, process writing, and writing strategy instruction. We found that implementing the program not only improved students' writing performance, but also increased teachers' writing strategy practice. However, we found no significant effect of using the writing program on communicative writing or process writing practices. Nor did we find a significant effect on the extent to which teachers taught learning strategies, differentiated, promoted active learning or on students' time on task. Students' time on task and the extent to which teachers promoted active learning were already high at the pretest, and therefore hard to improve. That teachers who implemented the comprehensive writing program did not differentiate significantly more often at the posttest than teachers who taught regular writing instruction could be due to the fact that implementing a new program is demanding for teachers, while differentiating is one of the most difficult teaching skills to master (Van de Grift, 2014; Kyriakides, Creemers, & Antoniou, 2009). To differentiate while implementing a new program is probably too demanding for most teachers.

Because the writing program is a comprehensive program and we studied it as a whole, we cannot determine which components contributed most to its effectiveness. However, because of the significant increase in teachers' writing strategy practice after implementing the program and as there were no significant changes in teachers' use of other classroom practices, it seems likely that writing strategy instruction contributed to some extent to the improvement in students' writing performance.[2] Indeed, we found a significant correlation between teachers' writing strategy practice and students' writing performance[2] at the posttest: (r = .38, p < .05). In addition, prior studies have

abundantly shown the effectiveness of writing strategy instruction (e.g. Graham & Perin, 2007; Graham et al., 2012; Koster et al., 2015).

There was a tendency that teachers who implemented the comprehensive writing program reported being significantly more efficacious in teaching strategies at the posttest than at the pretest. This correspondence between teachers' strategy-practice and their efficacy belief in teaching strategies is in line with several studies which showed that teachers' beliefs and writing classroom practices are related (e.g., Gaitas & Alves Martins, 2015; Lipson, Mosenthal, Daniels, & Woodside-Jiron, 2000; Troia, Lin, Cohen, & Monroe, 2011).

Contrary to our expectations, we did not find a direct additional effect of the professional development program on any of the outcome variables. There was no reason to assume, however, that the duration or intensity of the professional development program was insufficient. Our program met the criteria for professional development activities suggested by Desimone (2009), based on earlier studies, of at least 20 hours of interaction spread over a semester. Neither were there any indications that the program was of insufficient quality. As we reported in section 4.1, teachers were satisfied with the professional development program, and the majority of the teachers said that they learned a lot from the sessions.

The fact that we did not find any direct additional effect of the complementary professional development program on the outcome measures, might be due to the number of cases as only 11 teachers participated in the PD condition. This may have decreased the power to observe significant effects of the PD program, although at the student level the power was large. However, there is an indirect effect of the PD condition. After all, participating in the professional development program did significantly increase the number of writing program lessons teachers taught, compared to the writing program only condition. On average teachers in the PD condition taught six more lessons (three units), 29%, and we reported an effect of .71 points per unit taught, which equaled about one tenth of a grade effect. This effect might be due to the PD course itself: it might have been easier for the teachers in the PD condition to implement the lessons because they had received extensive support on how to do so, whereas the teachers in the WP condition were given less support (kick-off session, teacher manual). Another explanation is a possible stronger commitment of these teachers. Desimone (2009) argued that many studies have shown that teachers were more committed to innovation as a result of professional development programs. That might have been the effect of our PD condition: a stronger commitment, because of the regular interaction with the trainers, researchers and other teachers, which stimulated teachers to teach more writing lessons. That six teachers who were initially assigned to the PD condition were transferred to the WP condition at their request, may also indicate this stronger commitment of the teachers in the PD condition.

5.1 Strengths

We propose that this study has four strengths. First, we created a sustainable innovation. The writing lessons are embedded in a widely used online reading comprehension program and continue to be designed and provided every two weeks, with the opportunity for the designers to continuously improve the writing lessons' design. Furthermore, a spin-off of the professional development program is still offered to teachers, in addition to the lessons. So, teachers in the Netherlands can still profit from the writing program and the professional development program (in a reduced form) that were designed in this study.

Second, we attempted - although without direct success - to examine both the effectiveness of a writing program and the added value of a complementary professional development program. Other writing strategy intervention studies either tested the effectiveness of a writing program (e.g., Bouwer et al., 2016) or of a professional development program linked to a strategy-based writing program (e.g., Festas, Oliveira, Rebelo et al., 2015; Harris, Lane, Graham, et al., 2012; Harris, Graham, & Adkins, 2015; Koster et al., 2017).

A third strength is the generalizability of the findings, in line with Bouwer et al. (2016). Measuring students' writing performance with several tasks in different genres at each measurement occasion results in effect sizes that can be generalized over classes, students, and writing tasks. Estimates of effect sizes are therefore more realistic.

Fourth, this study examined the effects of an intervention on students' writing performance, teachers' classroom practices, and teachers' beliefs and skills. Previous strategy intervention studies only studied the effects on students' writing performance. Since students' writing performance can be improved by changing teachers' classroom practices, and since teachers' classroom practices and teachers' beliefs and skills are related (e.g., Rietdijk et al., in press; Gaitas & Alves Martins, 2015; Lipson, Mosenthal, Daniels, & Woodside-Jiron, 2000; Troia, Lin, Cohen, & Monroe, 2011), studying the effects of an intervention on teachers' classroom practices, beliefs and skills in addition to the effect on students' writing has added value, because it can shed light on what contributed to the intervention's effectiveness.

5.2 Future research

Now that we know that each additional lesson of the comprehensive writing program contributed to its effectiveness, a key issue is to determine what happened in these lessons that helped students learn to improve their writing more than in control classes? More detailed studies about what happens during learning, related to outcomes in terms of process and performance, would be welcome. We already know quite a lot about writing processes, but much less about learning-to-write processes. Additional output measures would also be informative. For instance, we now know that students learnt to write better texts than their peers in the control condition. We also know that teachers taught writing strategies more often than at the start of the intervention. More

insight in what students know about strategies, and whether and how they use these strategies could have been relevant outcome measures.

It was surprising that the professional development program, in addition to the lesson materials, did not affect classroom practice, other than increasing the number of lessons taught. A future study could investigate the effect of the professional development program on the quality of teachers' classroom practices. In the present study, we measured the presence of instructional elements for communicative, process-oriented and strategy-oriented writing education at pretest and posttest. Frequency of types of instructional behavior is the lowest level of instructional effectiveness (Kyriakides et al., 2009). In their dynamic model of educational effectiveness, Kyriakides et al. (2009) stress the importance of examining not only the quantity but also the quality of teaching activities. The number of times an activity takes place shows how important the teacher finds it, but disregards the fact that the way it functions may vary (Kyriakides et al., 2009). Applied to the present study: we might have observed the same amount of instructional behaviors, but they might have been better paced, integrated, and structured by the teachers in the PD condition.

Lastly, the current study examined the effects after one school year. We suggest that a future study investigates the effect of teachers using the writing program for a longer period of time. We expect that the comprehensive writing program is more effective when teachers use it for the second year because they will be familiar with its structure and principles, and it is likely that they will have improved their execution of the required skills (e.g., modeling of writing strategies, differentiating). Furthermore, they might need less preparation time, which could lead to an increase in the number of lessons they teach. At the same time, the program has lost its novelty when it is used for the second year, which might cause less investment in lesson preparation and lower quality of instructions as a result. This loss of novelty also holds for the student level: when students move on to a higher grade, and are offered the comprehensive writing program again, we may expect greater ease and consolidation of strategies due to the repeated instruction and more practice at one hand, but it might also be less motivating for students, and for the better writers, that they find themselves less challenged.

5.3 Conclusions

This study showed that a comprehensive writing program, combining communicative writing, process writing, and writing strategy instruction, improved students' writing performance and increased teachers' writing strategy practice tended to raise teachers' efficacy belief in teaching learning strategies in Grades 4 to 6 of Dutch primary schools. The program's effectiveness was influenced by the number of lessons taught; students' writing performance increased as additional lessons were implemented.

Direct additional effects of a complementary professional development program on students' writing and teachers' classroom practices were not found, but the teachers in the PD condition taught significantly more writing program lessons than the teachers in

the WP condition. So, indirectly the professional development program increased students' writing performance.

Notes

1. A combination of grades is quite common in Dutch elementary schools. These combination classes are usually formed because of the small number of students per grade, or for pedagogical reasons.
2. We calculated a residual score per teacher, i.e. the extent to which a teacher deviated from the mean.

Acknowledgements

References

Aarnoutse, C., & Kapinga, T. (2006). *Begrijpend Lezen 345678* [Reading Comprehension in grades 1 – 6]. Ridderkerk: 678 Onderwijs Advisering.

Basturkmen, H. (2012). Review of research into the correspondence between language teachers' stated beliefs and practices. *System*, *40*, 282-295. doi:10.1016/j.system.2012.05.001

Blok, H. (1986). Essay rating by the comparison method. *Tijdschrift voor Onderwijsresearch*, *11*(4),169–176.

Bochardt, I. (1984). Het schrijfproces: Cognitief-psychologisch onderzoek van Flower and Hayes. [The writing process; Cognitive-psychological research of Flower and Hayes.] *Tijdschrift voor taalbeheersing*, *6*(1), 23-42.

Bonset, H., & Hoogeveen, M. (2007). *Schrijven in het basisonderwijs: een inventarisatie van empirisch onderzoek in het perspectief van leerplanontwikkeling*. [Writing in primary education: A review of empirical research in the perspective of curriculum development]. Enschede: SLO.

Borg, S. (2009). Introducing language teacher cognition. Retrieved from http://www.education.leeds.ac.uk/assets/files/staff/borg/Introducing-language-teacher-cogni–tion.pdf

Bouwer, R., Béguin, A., Sanders, T., & Van den Bergh, H. (2014). Effect of genre on the generalizability of writing scores. *Language Testing, 32*(1), 83-100. doi: 10.1177/0265532214542994

Bouwer, R., & Koster, M. (2016). *Bringing writing research into the classroom: The effectiveness of Tekster, a newly developed writing program for elementary students*. (Doctoral dissertation). Utrecht: Universiteit van Utrecht.

Braet, A. (1979). *Taaldaden: Een leergang schriftelijke taalbeheersing*. [Speech acts: A curriculum on writing and reading]. Groningen: Wolters-Noordhoff.

Brindle, M., Graham, S., Harris, K. R., & Hebert, M. (2016). Third and fourth grade teachers'classroom practices in writing: A national survey. *Reading and Writing*, *29*, 929-954. doi:10.1007/s11145-015-9604-x

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

Drop, W., & De Vries, J. (1976). *Ter informatie: Leergang samenvatten en schrijven van zakelijke teksten*. [For your information. How to summarize and write informative texts.] Groningen: Wolters-Noordhoff.

Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. B. (2003). A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health Education Research, 18*(2), 237 – 256. doi:10.1093.her.18.2.237

Evers-Vermeul, J., & Van den Bergh, H. (2009). Schrijf voor de lezer: Over effecten van lezersgericht (her)schrijven op de kwaliteit van instructieve teksten. [Write for the reader: On the effects of reader oriented (re)writing on the quality of instructive texts]. *Levende Talen Tijdschrift*, *10*, 14-23.

Expert Group Learning Trajectories (2009). *Referentiekader taal en rekenen: De referentieniveaus*. [Reference framework language and arithmetic: The referential levels]. Enschede: Ministerie van OCW.

Festas, I., Oliveira, A. L., Rebelo, J. A., Damião, M. H., Harris, K., & Graham, S. (2015). Professional development in Self-Regulated Strategy Development: Effects on the writing performance of eighth grade Portuguese students. *Contemporary Educational Psychology*, 40, 17-27. doi: 10.1016/j.cedpsych.2014.05.004

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

Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication*, *32*, 365-387. doi:10.2307/356600.

Franssen, H. M. B., & Aarnoutse, A. (2003). Schrijfonderwijs in de praktijk. [Writing education in practice]. *Pedagogiek*, *23*(3), 185-198.

Gaitas, S., & Alves Martins, M. (2015). Relationships between primary teachers' beliefs and their practices in relation to writing instruction. *Research Papers in Education*, *30*(4), 492-505. doi:10.1080/ 02671522.2014.908406

Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, *38*(4), 915-945. doi: 10.3102/00028312038004915

Gibson, S., & Dembo, M. (1984). Teacher efficacy: A construct validation. *Journal of Educational Psychology*, *76*, 569–582. doi:10.1037/0022-0663.76.4.569

Graham, S., Harris, K.R., Fink, B., & MacArthur, C.A. (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. R., MacArthur, C.A., & Fink, B. (2002). Primary grade teachers' theoretical orientations concerning writing instruction: Construct validation and a nationwide survey. *Contemporary Educational Psychology, 27*, 147-166. doi:10.1006/ceps.2001.1085

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

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

Graham, S., & Sandmel, K. (2011). The process writing approach: A meta-analysis. *The Journal of Educational Research*, *104*(6), 396-407. doi: 10.1080/00220671.2010.488703

Harn, B., Parisi, D., & Stoolmiller, M. (2013). Balancing Fidelity with Flexibility and Fit: What do we really know about Fidelity of Implementation in Schools? *Exceptional Children, 79*(2), 181 – 193. doi: 10.1177/00144029130790024

Harris, K. R., & Graham, S. (2009). Self-regulated strategy development in writing: Premises, evolution, and the future. In *BJEP Monograph Series II, Number 6   Teaching   and   Learning Writing* (Vol. 113, No. 135, pp. 113-135). British Psychological Society.

Harris, K. R., Graham, S., & Adkins, M. (2015). Practice-based professional development   and self-regulated strategy development for Tier 2, at-risk writers in second grade.   *Contemporary Educational Psychology*, *40*, 5-16. doi:10.1016/j.cedpsych.2014.02.003

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*(2), 103-119. doi: 10.1177/0022487111429005

Hayes, J.R. (1996). A New Framework for Understanding Cognition an Affect in Writing. In: C.M. Levy, S. Randell (red.). *The Science of Writing. Theories, Methods, Individual Differences and Applications*. Mahwah, New Jersey: Lawrence Erlbaum.

Henk, W. A., Marinak, B. A., Moore, J. C., & Mallette, M. H. (2003). The writing observation framework: A guide for refining and validating writing instruction. *The Reading Teacher*, *57*, 322-333.

Henkens, L. (2010). *Het onderwijs in het schrijven van teksten. De kwaliteit van het schrijfonderwijs in het basisonderwijs*. [Writing education: The quality of writing education in primary schools]. Utrecht: Inspectie van het Onderwijs.

Holliway, D., & McCutchen, D. (2004). Audience perspective in children's descriptive writing: Reading as the reader. In G. Rijlaarsdam (Series Ed) & L. Allal, L. Chanquoy, & P. Largy (Vol. Eds.), *Revision of written language: Cognitive and instructional processes. Studies in Writing*, vol. 13 (pp. 87-101). Springer Netherlands. doi:10.1007/978-94-007-1048-1

Hoogeveen, M., & Van Gelderen, A. (2013). What works in writing with peer response? A review of intervention studies with children and adolescents. *Educational Psychology Review*, *25*(4), 473-502. doi: 10.1007/s10648-013-9229-z

Hymes, Dell H. (1972). On communicative competence. In J. B. Pride & J. Holmes (Eds.), *Sociolinguistics: selected readings* (pp. 269-293). Harmondsworth: Penguin.

Karweit, N., & Slavin, R. E. (1982). Time-on-task: Issues of timing, sampling, and definition. *Journal   of   Educational   Psychology*, *74*(6),   844-851.   doi: 10.1037/0022-0663.74.6.844

Kirkpatrick, L.C., & Klein, P.D. (2009). Planning text structure as a way to improve students' writing from sources in the compare-contrast genre. *Learning and Instruction*, *19*, 309-321. doi: 10.1016/j.learninstruc.2008.06.001

Koster, M., Bouwer, R., & Van den Bergh, H. (2017). Professional development of teachers in the implementation of a strategy-focused writing intervention program for elementary students. *Contemporary Educational Psychology*, *49*, 1-20. doi: 10.1016/j.cedpsych.2016.10.002

Koster, M., Tribushinina, E., Jong, P. de, & Van den Bergh, H. (2015). Teaching children to write: A meta-analysis of writing intervention research. *Journal of Writing Research, 7* (2), 300-324. doi: 10.17239/jowr-2015.07.02.2

Kretlow, A. G., & Bartholomew, C. C. (2010). Using coaching to improve the fidelity of evidence-based practices: A review of studies. *Teacher Education and Special Education: The Journal of the Teacher Education Division of the Council for Exceptional Children, 33*(4), 279-299. doi: 10.1177/0888406410371643

Krom, R., Van de Gein, J., Van der Hoeven, J., Van der Schoot, F., Verhelst, N., Veldhuijzen, N., & Hemker, B. (2004). *Balans van het schrijfonderwijs op de basisschool. Uitkomsten van de peilingen in 1999: halverwege en eind basisonderwijs en speciaal basisonderwijs.* [Evaluation of writing skills in primary education and special education: Results of the assessments in 1999]. Arnhem: Stichting Cito Instituut voor Toetsontwikkeling.

Kroon, J. W. M. (1985). *Grammatica en communicatie in het onderwijs Nederlands.* [Grammar and communication in Dutch language education]. (Doctoral dissertation). Groningen: Wolters-Noordhoff.

Kuhlemeier, H., Van Til, A., Hemker, B., De Klijn, W., & Feenstra, H. (2013). *Balans van de schrijfvaardigheid in het basis en speciaal onderwijs 2.* [Second evaluation of writing skills in primary education and special education]. Arnhem: Stichting Cito Instituut voor Toetsontwikkeling.

Kyriakides, L., Creemers, B.P.M., & Antoniou, P. (2009). Teacher behaviour and student outcomes: Suggestions for research on teacher training and professional development. *Teaching and Teacher Education, 25,* 12-23. doi:10.1016/j.tate.2008.06.001

Leidse Werkgroep Moedertaaldidactiek (1980). *Moedertaaldidactiek. Een handleiding voor het voortgezet onderwijs.* [Mother tongue teaching. A guide for secondary education]. Muiderberg: Coutinho.

Lipson, M.Y., Mosenthal, J., Daniels, P., & Woodside-Jiron, H. (2000). Process writing in the classrooms of eleven fifth-grade teachers with different orientations to teaching and learning. *The Elementary School Journal, 101*(2), 209 – 231. doi: 10.1086/499665

McKeown, D., Brindle, M., Harris, K. R., Graham, S., Collins, A. A., & Brown, M. (2016). Illuminating growth and struggles using mixed methods: Practice-based professional development and coaching for differentiating SRSD instruction in writing. *Reading and Writing, 29,* 1105-1140. doi: 10.1007/s11145-016-9627-y

Merrill, M.D. (2002). First Principles of Instruction. *Educational Technology Research and Development, 50*(3), 43-59. doi:10.1007/BF02505024

Mowbray, C. T., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity Criteria: Development, Measurement, and Validation. *American Journal of Evaluation, 24*(3), 315 – 340. doi: 10.1177/109821400302400303

O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K-12 curriculum intervention research. *Review of Educational Research, 78,* 33–84. doi:10.3102/0034654307313793

Opfer, V. D., & Pedder, D. (2011). The lost promise of teacher professional development in England. *European Journal of Teacher Education, 34,* 3–24. doi: 10.1080/02619768.2010.534131

Penuel, W. R., Fishman, B. J., Yamaguchi, R., & Gallagher, L. P. (2007). What makes professional development effective? Strategies that foster curriculum implementation. *American Educational Research Journal, 44*(4), 921-958. doi: 10.3102/0002831207308221

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] (Doctoral dissertation). Utrecht: Universiteit van Utrecht.

Rietdijk, S., Van Weijen, D., Janssen, T., Van den Bergh, H., & Rijlaarsdam. G. (in press). Teaching Writing in Primary Education: Classroom Practices, Learning Time, and Teachers' Beliefs. *Journal of Educational Psychology.*

Rijlaarsdam, G. (1986). *Effecten van leerlingenrespons op aspecten van stelvaardigheid.* [Effects of student peer feedback on aspects of written composition skills]. (Doctoral dissertation). Amsterdam: SCO.

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

Rijlaarsdam, G., Braaksma, M., Couzijn, M., Janssen, T., Kieft, M., Raedts, M., Steendam, E. van, Toorenaar, A., & Van den Bergh, H. (2009). The role of readers in writing development:

Writing students bringing their texts to the test. In R. Beard, D. Myhill, J. Riley & M. Nystrand (Eds.), *The SAGE handbook of writing development* (pp.436-452). London: SAGE Publications. doi: 10.4135/9780857021069.n31

Schoonen, R. (2012). The validity and generalizability of writing scores: the effect of rater, task and language. In Van Steendam, E., Tillema, M., Rijlaarsdam, G., & Van den Bergh, H. (Eds.), *Measuring writing: recent insights into theory, methodology and practices* (Studies in writing, 27) (pp. 1-22). Leiden-Boston: Brill.

Schoonen, R., & De Glopper, K. (1996). Writing performance and knowledge about writing. In G. Rijlaarsdam, H. van den Bergh & M. Couzijn (Eds.), *Theories, models & methodology in writing research* (pp. 87-107). Amsterdam: Amsterdam University Press.

Snijders, T., & Bosker, R. (1999). *Multilevel modeling: An introduction to basic and advanced multilevel modeling.* Sage, Thousand Oaks, CA.

Ten Brinke, S. (1976). *The complete mother-tongue curriculum.* Groningen: Wolters-Noordhoff-Longmann.

Troia, G.A., & Maddox, M.E. (2004). Writing instruction in middle schools: Special and general education teachers share their views and voice their concerns. *Exceptionality, 12*(1), 19-37. doi: 10.1207/s15327035ex1201_3

Troia, G.A., Lin, S.C., Cohen, S., & Moore, B.W. (2011). A Year in the Writing Workshop: Linking Writing Instruction Practices and Teachers' Epistemologies and Beliefs about Writing Instruction. *The Elementary School Journal, 112*(1), 155-182. doi: 10.1086/660688

Van den Bergh, H., & Eiting, M. H. (1989). A method of estimating rater reliability. *Journal of Educational Measurement, 26*(1), 29-40. doi: 10.1111/j.1745-3984.1989.tb00316.x

Van de Grift, W. (2007). Quality of teaching in four European countries: A review of the literature and application of an assessment instrument. *Educational Research, 49*(2), 127-152, doi: 10.1080/00131880701369651

Van de Grift, W. (2014). Measuring teaching quality in several European countries. *School Effectiveness and School Improvement, 25*(3), 295-311. doi: 10.1080/ 09243453.2013.794845

Van der Leeuw, B. (2006). *Schrijftaken in de lerarenopleiding: Een etnografie van onderwijsvernieuwing.* [Writing tasks in teacher training: An etnography of educational innovation]. (Doctoral dissertation). Utrecht: Universiteit van Utrecht.

Van Veen, K., Zwart, R., Meirink, J., & Verloop, N. (2010). *Professionele ontwikkeling van leraren. Een reviewstudie naar effectieve kenmerken van professionaliserings-interventies van leraren.* [Professional development of teachers: A review of effective professional development programs.] Leiden: ICLON.

Wesdorp, H. (1974). *Het meten van productieve-schriftelijke taalvaardigheid* [Measurement of written language skills]. (Doctoral dissertation). Purmerend: The Netherlands: Muusses.

White, M.J., & Bruning, R. (2005). Implicit writing beliefs and their relation to writing quality. *Contemporary Educational Psychology, 30*, 166-189. doi: 10.1016/j.cedpsych.2004.07.002

Zwarts, M., Rijlaarsdam, G. C. W., Janssens, F., Wolfhagen, I., Veldhuijzen, N., & Wesdorp, H. (1990). *Balans van het taalonderwijs aan het einde van de basisschool.* [*Survey of language assessment at the end of primary education.*] Arnhem: Cito Instituut voor Toetsontwikkeling.

Appendix A: Example of a genre-specific writing strategy

A strategy for writing descriptive texts
A descriptive text states what an object, person or place looks like. The text contains a description of parts or qualities, but should also provide an overall impression of the object, person or place described. The communicative goal of descriptive texts is to inform.

We developed the BEVERS strategy (Dutch for BEAVERS) to help students write good descriptions. This strategy is a modified version of the IAPN planning strategy developed by Kirkpatrick and Klein (2009) for writing compare-contrast reports, in which IAPN stands for Information, Aspect, Paragraph, and Number (or sequence of the information in the text).
The strategy steps were:

1. <u>Observe</u> the object: what strikes you the most? Make a note of this (BE)
2. <u>Compare</u> the object to other objects of the same kind: in what ways is this object different? Write down the differences (VER).
3. Order your findings in a <u>schematic outline</u> (S).

Example of a schematic outline:

What I noticed about the object:	*This says something about*	*This will be the order in my text*
The object is white	Colour	3
It is much larger than other objects	Size	1
It is round	Shape	2
….	…..	4

Students subsequently write their text on the basis of their schematic outline.

Appendix B: Reliability of the questionnaires

	Cronbach's alpha (α)		
	Pretest	Posttest	Number of items deleted
Beliefs about writing			
- Writing as transmission	.69	.70	4
- Writing as transaction	.78	.70	3
Beliefs about writing instruction			
- Correct writing	.68	.68	1
- Explicit instruction	.66	.63	2
- Natural learning	.61	.68	3
Efficacy beliefs in teaching writing			
- Personal teaching efficacy	.62	.67	4
- General teaching efficacy	.71	.64	2
Efficacy in providing high-quality instruction			
- Efficacy in teaching learning strategies	.83	.86	1
- Efficacy in differentiating	.88	.88	0
- Efficacy in promoting active learning	.85	.76	4
Providing high-quality instruction			
- Teaching learning strategies	.85	.81	0
- Differentiating	.87	.85	0
- Promoting active learning	.84	.80	0

Appendix C: Statistics

Table C1: *Significance Values of Communicative Writing Analyses*

Variable	df	F	p
Goal directedness is mentioned while discussing text quality			
measurement occasion	1, 82	17.53	<.001*
condition	1, 82	2.37	.13
condition x measurement occasion	1, 82	.01	.94
Feedback on goal directedness			
measurement occasion	1, 75	2.14	.15
condition	1, 75	.23	.63
condition x measurement occasion	1, 75	.02	.88
Audience awareness is mentioned while discussing text quality			
measurement occasion	1, 82	.87	.35
condition	1, 82	.18	.67
condition x measurement occasion	1, 82	.87	.35
Feedback on audience awareness			
measurement occasion	1, 75	3.74	.06
condition	1, 75	.01	.91
condition x measurement occasion	1, 75	.02	.89
Students' texts are read aloud			
measurement occasion	1, 82	1.30	.26
condition	1, 82	1.80	.18
condition x measurement occasion	1, 82	1.30	.26
Students' texts are published			
measurement occasion	1, 75	.10	.75
condition	1, 75	.01	.94
condition x measurement occasion	1, 75	.88	.35

* significant at the 0.05 level

Table C2: Significance Values of Process Writing Analyses

Variable	df	F	p
Generating ideas			
measurement occasion	1, 81	.08	.13
condition	1, 81	2.40	.77
condition x measurement occasion	1, 81	.002	.97
Organizing ideas			
measurement occasion	1, 81	.002	.96
condition	1, 81	1.74	.19
condition x measurement occasion	1, 81	.60	.44
Revising texts			
measurement occasion	1, 74	.01	.92
condition	1, 74	3.65	.06
condition x measurement occasion	1, 74	.03	.87

Table C3: Significance Values of Writing Strategy Instruction Analyses

Variable	df	F	p
Teaching writing strategies			
measurement occasion	1, 81	1.33	.25
condition	1, 81	4.43	.04*
condition x measurement occasion	1, 81	14.05	.001*
Modeling the writing process			
measurement occasion	1, 81	1.53	.22
condition	1, 81	.14	.71
condition x measurement occasion	1, 81	.84	.36

* significant at the 0.05 level

Table C4: Significance Values of High Quality Instruction Analyses

Variable	*df*	*F*	*p*
Teaching learning strategies			
measurement occasion	1, 41	8.88	.01*
condition	1, 42	4.17	.05*
condition x measurement occasion	1, 41	.85	.36
Differentiating			
measurement occasion	1, 42	.58	.45
condition	1, 43	.004	.95
condition x measurement occasion	1, 42	1.34	.25
Promoting active learning			
measurement occasion	1, 43	5.43	.03*
condition	1, 43	6.05	.02*
condition x measurement occasion	1, 43	.93	.34

* Significant at the 0.05 level

Table C5: Significance Values of Teacher Beliefs Analyses

Variable	df	F	p
Writing as transmission			
measurement occasion	1, 41	.02	.90
condition	1, 42	.003	.95
condition x measurement occasion	1, 41	1.75	.19
Writing as transaction			
measurement occasion	1, 42	1.37	.35
condition	1, 43	.90	.25
condition x measurement occasion	1, 42	.72	.40
Correct writing			
measurement occasion	1, 41	.02	.88
condition	1, 42	1.30	.26
condition x measurement occasion	1, 41	3.80	.06
Explicit instruction			
measurement occasion	1, 43	1.08	.31
condition	1, 44	.12	.73
condition x measurement occasion	1, 43	.69	.41
Natural learning			
measurement occasion	1, 43	2.62	.11
condition	1, 44	3.36	.07
condition x measurement occasion	1, 43	2.36	.13
Personal teaching efficacy			
measurement occasion	1, 43	2.46	.12
condition	1, 44	.05	.82
condition x measurement occasion	1, 43	.84	.36
General teaching efficacy			
measurement occasion	1, 41	.16	.69
condition	1, 42	1.02	.32
condition x measurement occasion	1, 41	.09	.77
Efficacy in teaching learning strategies			
measurement occasion	1, 42	19.28	<.001*
condition	1, 43	5.10	.03*
condition x measurement occasion	1, 42	4.45	.04*
Efficacy in differentiating			
measurement occasion	1, 43	2.41	.13
condition	1, 44	.19	.67
condition x measurement occasion	1, 43	1.34	.25
Efficacy in promoting active learning			
measurement occasion	1, 42	16.96	<.001*
condition	1, 43	1.33	.26
condition x measurement occasion	1, 42	.75	.39

Table C6: Significance Values of Teacher Skills Analyses

Variable	df	F	p
Text assessment			
measurement occasion	1, 38	.50	.49
condition	2, 39	.84	.44
condition x measurement occasion	2, 38	2.56	.09
Reflecting on lessons			
measurement occasion	2, 71	2.89	.06
condition	2, 37	.61	.55
condition x measurement occasion	4, 71	1.90	.12
Adapting of lessons to context			
measurement occasion	2, 68	.06	.95
condition	2, 33	.46	.64
condition x measurement occasion	4, 68	.26	.90